(Continued)

CONTENTS—*Continued from preceding page*

CONTENTS—*Continued from preceding page*

CONTENTS—*Continued from preceding page*

Characterization of uncertainty in outdoor sound propagation predictions[a]

D. Keith Wilson, Edgar L Andreas, and John W. Weatherly
U.S. Army Engineer Research and Development Center, 72 Lyme Rd., Hanover, New Hampshire 03755
d.keith.wilson@erdc.usace.army.mil, edgar.l.andreas@erdc.usace.army.mil, john.w.weatherly@erdc.usace.army.mil

Chris L. Pettit
Aerospace Engineering Department, U.S. Naval Academy, 590 Holloway Rd., MS-11B,
Annapolis, Maryland 21402
petitcl@usna.edu

Edward G. Patton and Peter P. Sullivan
National Center for Atmospheric Research, P.O. Box 3000, Boulder, Colorado 80307
patton@ucar.edu, pps@ucar.edu

Abstract: Predictive skill for outdoor sound propagation is assessed using high-resolution atmospheric fields from large-eddy simulations (LES). Propagation calculations through the full LES fields are compared to calculations through subsets of the LES fields that have been processed in typical ways, such as mean vertical profiles and instantaneous vertical profiles synchronized to the sound propagation. It is found that mean sound pressure levels can be predicted with low errors from the mean profiles, except in refractive shadow regions. Prediction of sound pressure levels for short-duration events is much less accurate, with errors of $8-10$ dB for near-ground propagation being typical.

1. Introduction

Propagating sound waves are sensitive to complex, four-dimensional details of the atmosphere that cannot be fully resolved with available observation systems. This insufficiency in the data introduces uncertainty into sound propagation predictions. For example, a common situation is to have available wind and temperature data collected at only a few discrete heights on a single, vertical tower, in which case the vertical structure of the atmosphere is coarsely sampled, while the horizontal structure is not directly sampled at all. Additionally, the location where the atmospheric data are collected often does not coincide with the path of sound propagation. Numerical weather forecast data, which are becoming increasingly used, sample all three spatial dimensions and time but at a much lower resolution than needed for accurate predictions. Given the practical compromises involved, it is necessary to assess the skill of propagation predictions for particular types and resolutions of atmospheric inputs.

The CASES '99 experiment, described in Ref. 1, illustrates how predictive skill is often limited by the atmospheric data. This experiment demonstrated that, even on a clear night over flat terrain with wind and temperature collected at 5 m increments on a 50 m tower, the timing and location of $5-20$ dB fluctuations in sound level could not be consistently predicted.

This study uses a simulation approach to assess predictive uncertainty for sound propagation at low frequencies. Turbulence simulations are performed at as high a resolution as

---

[a]Preliminary results from this study were presented in D. K. Wilson, E. L Andreas, J. W. Weatherly, and C. L. Pettit, "Uncertainty in outdoor sound propagation predictions as determined from high-resolution atmospheric simulations," InterNoise 2006, Honolulu, HI.

Table 1. Parameters characterizing the four large-eddy simulations used in this study.

Stratification	Very unstable	Unstable	Neutral	Stable
Friction velocity (m s^{-1})	0.317	0.295	0.219	0.284
Roughness length (m)	0.02	0.02	0.02	0.1
Sensible heat flux (W m^{-2})	115	29	0	−13.3
Geostrophic wind speed (m s^{-1})	5.0	5.0	5.0	8.0
Boundary-layer height (m)	650	650	615	195
Dimensions (km)	$2.4 \times 2.4 \times 1$	$2.4 \times 2.4 \times 1$	$2.4 \times 2.4 \times 1$	$1 \times 1 \times 0.4$
Resolution (m)	$4 \times 4 \times 2.5$	$4 \times 4 \times 2.5$	$4 \times 4 \times 2.5$	$2 \times 2 \times 1$
Total duration (s)	983	914	710	653
Number of saved volumes	200	200	200	120

practical to create surrogate atmospheric fields. Sound is then propagated through these fields, and the results are statistically compared to predictions based on less detailed but more typical atmospheric data sets, such as mean vertical profiles and instantaneous vertical profiles at a location near the propagation path.

2. Description of the atmospheric simulations

Our atmospheric simulations of the wind and temperature fields are based on the method of large-eddy simulation (LES), which is commonly applied to high Reynolds number flows such as the atmosphere.[2] The simulations were performed on a parallel-processing supercomputer with 100 CPUs in order to obtain the highest practical resolution. Four different atmospheric density stratifications were simulated; these are termed *very unstable*, *unstable*, *neutral*, and *stable*. Unstable stratification is characteristic of sunny days, neutral stratification of windy and cloudy conditions, and stable stratification of clear nights. The ground is flat with uniform roughness elements. The important parameters for the four runs are shown in Table 1.

Mm. 1 and Mm. 2 show vertical cross sections, aligned with the nominal (geostrophic) wind direction, for the very unstable and stable stratification. In unstable stratification, plumes of relatively warm air rise from the heated surface. The air 50 m or so above the surface tends to be well mixed. In stable stratification, wave motions form and there is much less mixing. The animations also show synchronized upwind and downwind sound propagation simulations, which were performed by the method to be described in the next section.

> Mm 1. Turbulence fields and sound propagation in very unstable stratification (10.2 Mb). This is a file of type "avi." The animation spans about 16.4 min. Top: vertical cross sections through the LES. Coloration indicates potential temperature in C, vectors indicate wind. Middle: Downwind propagation at 150 Hz through the LES fields as calculated by the parabolic equation method. Coloration represents sound level relative to 0 dB at 1 m from the source. Source is on the left side of the image, at 1 m height. Bottom: upwind propagation. Source is on the right side of the image, at 1 m height.

> Mm. 2. Turbulence fields and sound propagation in stable stratification (6.0 Mb). This is a file of type avi. The animation spans about 10.9 min. See caption to Mm. 1 for additional information.

3. Sound propagation calculations and procedure

The sound-propagation calculations were performed with a two-dimensional, wide-angle, Crank-Nicholson, parabolic equation (CNPE) code of the type described by West *et al.*,[3] which is accurate for propagation angles within 20–30° of horizontal.[4] Height- and range-dependent cross sections of the LES fields can be readily incorporated into the CNPE, which takes as input the sound speed (which depends on temperature) and the horizontal wind component in the direction of propagation. The source height in all calculations is set to 1 m, and the ground properties are characteristic of short grass, as described in Ref. 1. Predictions in the nominal

downwind, upwind, and crosswind directions, out to a range of 1 km, were calculated at the frequencies 50, 150, and 250 Hz. CNPE calculations were repeated for each available saved volume (snapshot) from the LES simulations.

In discussing skill of the sound propagation predictions, *event* and *mean* predictions are distinguished and separately examined here. A propagation event has short duration compared to the time scales (associated with turbulence, internal waves, etc.) over which the atmosphere substantially modifies the propagation. Event prediction could apply to propagation of an explosion, or to predicting the sound level produced by a steady source over a short duration. Mean prediction involves predicting the ensemble mean-square sound pressure for a particular set of stationary atmospheric forcings. Mean-square sound pressure predictions over time intervals in which the forcings change, such as daily or seasonal intervals, would represent another distinct case for a predictive skill analysis.

Similarly, atmospheric data may be collected and processed on an event (equivalently, *snapshot* or *sample*) or mean basis. For example, a single radiosonde (weather balloon) ascent provides a snapshot of the atmosphere along a particular time/space trajectory. Forecasts from numerical weather prediction (NWP) models, on the other hand, typically provide ensemble mean (Reynolds averaged) atmospheric fields. Although it might seem natural that event propagation should be predicted with event atmospheric data and mean propagation predicted with mean atmospheric data, there is no consistent practice for doing so. For example, mean sound levels are often predicted from radiosonde profiles (event data), and NWP results (mean data) have been used to predict propagation of explosions.[5]

In the following, we evaluate the accuracy of event and mean propagation predictions using event and mean atmospheric data. Specifically, we consider the following types of processed atmospheric data:

- *Ensemble-mean vertical profiles*, which are calculated by averaging over horizontal planes in each LES volume, and then over time.
- *Instantaneous, along-path mean vertical profiles*, which are averages of the vertical profiles along the propagation path at the same time as the propagation event. Hence these are spatial, but not temporal, averages.
- *Instantaneous, midpoint vertical profiles* collected at a distance of 500 m from the source along the direction of propagation. When predicting event propagation, they are synchronized to the event. These profiles are at the true midpoint of the propagation path only for the 1000 m receiver distance.
- *Instantaneous, displaced vertical profiles* are the same as the instantaneous, midpoint profiles, except that the collection point is moved an additional 500 m in the direction transverse to the propagation (the positive crosswind direction).

The bias and mean-square errors associated with the various predictions are calculated from, respectively,

$$b = \frac{1}{N}\sum_{i=1}^{N} (\hat{\phi}_i - \phi_i), \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (\hat{\phi}_i - \phi_i)^2. \tag{1}$$

Here, i is the index of a particular LES volume snapshot, N is the number of snapshots, $\hat{\phi}_i$ is the value of the estimator associated with that snapshot (namely, the sound level prediction in dB), and ϕ_i is the actual value to be estimated, as determined from the fully resolved LES fields. In some cases, the $\hat{\phi}_i$ and/or ϕ_i may be the same for all of the snapshots; for example, if biases associated with predictions of the mean sound level are being calculated, the ϕ_i are all equal to the mean sound level in dB.
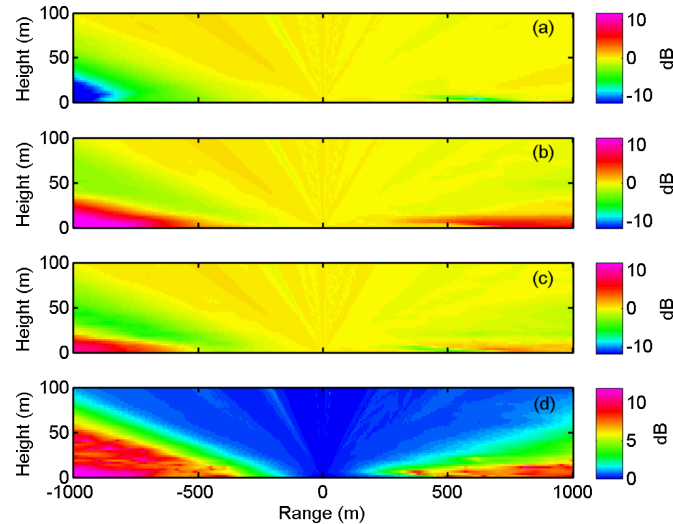
Fig. 1. (Color online) Errors for mean sound pressure level estimates at 150 Hz in unstable stratification. The left and right halves of the plot are upwind and downwind propagation, respectively, from a source located at 0 m range and 1 m height. (a) Bias error for estimation using ensemble mean vertical profiles. (b) Bias error for averaging propagation predictions from a large number of samples of the vertical profiles. (c) Bias error for estimation from a single, randomly collected set of vertical profiles. (d) Root mean square error corresponding to (c).

4. Results and discussion

4.1 Predictability of mean sound levels

We first consider prediction of mean sound levels from various types of atmospheric inputs. The actual mean levels (the ϕ_i in this case, which are the same for all i) are calculated by propagating the sound through each fully resolved LES cross section, averaging the mean-square sound pressures, and then converting to dB. Each image in Fig. 1 shows errors associated with a particular type of atmospheric input data, with the left side of each image representing upwind propagation and the right side downwind propagation. All calculations are for 150 Hz and unstable stratification. The upwind propagation exhibits strong, upward refraction. The downwind propagation exhibits weak ducting near the ground and upward refraction higher up. Errors tend to be largest near the ground, probably due to the greater importance of refraction at shallow propagation angles.

When ensemble-mean vertical profiles are used to form the estimate (the $\hat{\phi}_i$, which are the same in this case for all i) to the actual mean sound level [Fig. 1(a)], sound levels more than 12 dB too low occur near the ground in the upwind refractive shadow. A smaller under prediction of about 4 dB, which is associated with destructive interference between propagating modes, occurs downwind between approximately 400 and 750 m. These errors are attributable to the absence of turbulent scattering in calculations based solely on mean profiles.[6]

Figure 1(b) shows the bias error for mean-sound-level prediction based on calculating the mean-square sound pressure from the instantaneous, displaced vertical profile in each LES snapshot, averaging, and then converting to dB. An approach similar to this one was suggested by Yokota *et al.*,[7] who estimated mean sound levels by averaging predictions from a number of instantaneous, vertical profiles. In comparison to Fig. 1(a), the bias errors are similar in magnitude but opposite in sign. In effect, this estimation method assumes that the turbulent eddies have a spatial extent much longer than the propagation path, which leads to stronger scattering than actually occurs.

Figures 1(c) and 1(d) show the bias and root-mean-square [rms, σ from Eq. (1)] errors associated with estimating the mean sound level from a single random sample of instantaneous,
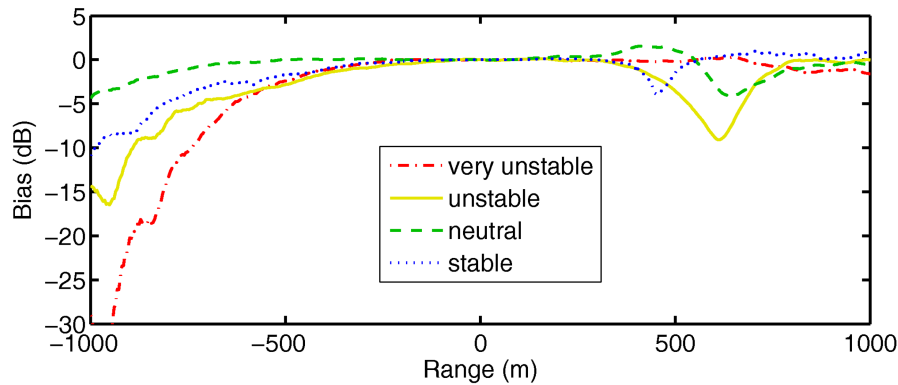
Fig. 2. (Color online) Bias errors for mean sound pressure level estimates at 150 Hz based on ensemble mean vertical profiles. Four different atmospheric stratifications are shown. Receiver height is 2 m.

displaced vertical profiles. That is, event atmospheric data are used to estimate mean sound level and $\hat{\phi}_i$ is therefore different for each LES snapshot i. The bias error is similar to Fig. 1(b), although there is a more pronounced negative bias along the upwind shadow zone boundary. There are very substantial rms errors, of 8 or more dB at distances greater than 400 m for near-ground propagation, in both the upwind and downwind directions.

  Figure 2 shows the bias error for estimates of the mean sound level using the ensemble mean vertical profiles. Calculations are shown for all four cases of the stratification, and the receiver height is fixed to 2 m. Under prediction of sound levels is found to be a problem for all stratifications but particularly so for upwind propagation in very unstable stratification, which creates the strongest refractive shadow.

4.2 Predictability of event sound levels

We next consider the accuracy of event sound-level predictions. In compiling the statistics, there is a separate random event (and hence a distinct ϕ_i) for each LES snapshot. Figure 3 shows rms errors for predictions based on each of the four profile types. All calculations are for 150 Hz and unstable stratification. We see that the along-path mean profiles actually produce the best predictions of event sound levels. They are slightly better than predictions based on the ensemble mean profiles. The predictions become much worse when instantaneous (event) profiles are used; the displaced profiles are only slightly worse than the midpoint ones. Apparently, there is enough fine detail in the atmospheric structure along the propagation path that observations at one location and time are not sufficiently representative to produce accurate predictions of a propagation event occurring at that same time. Typical prediction errors, for receivers near the ground at distances greater than 500 m, are $5-8$ dB when using the mean profiles and $8-10$ dB when using the instantaneous profiles.

  The bias errors corresponding to event predictions are not shown here. For the ensemble-mean predictions, the appearance is highly similar to Fig. 1(a). The along-path mean predictions have a similar pattern, but the bias is reduced to a few dB in the shadow region. The instantaneous profiles produce bias errors essentially the same as Fig. 1(c); that is, the scattering into shadow and interference zones is too strong.

  Figure 4 shows the effect of changing frequency. Calculations for downwind propagation in stable stratification, crosswind propagation in neutral stratification, and upwind propagation in unstable stratification are shown. The errors at 50 Hz are comparatively small and increase steadily with range. For downwind propagation in stable stratification, the 150 and 250 Hz cases both saturate at a similar error of about 8 dB. The saturation occurs at a shorter distance for the higher frequency. Although other stratifications are not plotted here, they all have essentially the same behavior in the downwind direction. For the crosswind propagation in
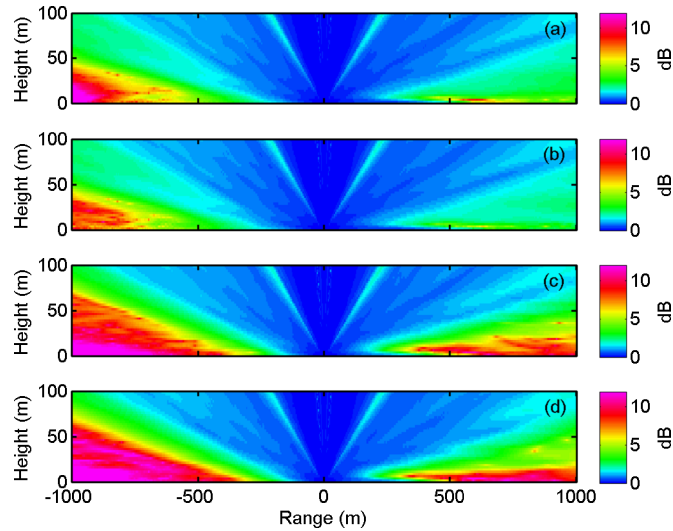
Fig. 3. (Color online) Root mean square errors for estimates of the event sound pressure levels at 150 Hz in unstable stratification by several different methods. (a) Estimation based on ensemble mean vertical profiles. (b) Estimation based on path-averaged vertical profiles. (c) Estimation based on instantaneous vertical profiles at midpoint of the propagation path. (d) Estimation based on instantaneous vertical profiles displaced from the propagation path.

neutral stratification and upwind propagation in unstable stratification, the errors tend to increase without saturation as the distance increases. This behavior is actually observed for all stratifications in the upwind and crosswind directions. The errors are considerably larger, up to about 18 dB, in the upwind direction.

5. Conclusions

This study has explored fundamental limitations in predictive skill for outdoor sound propagation. Such limitations are imposed by imperfect spatial and temporal sampling of the atmo-
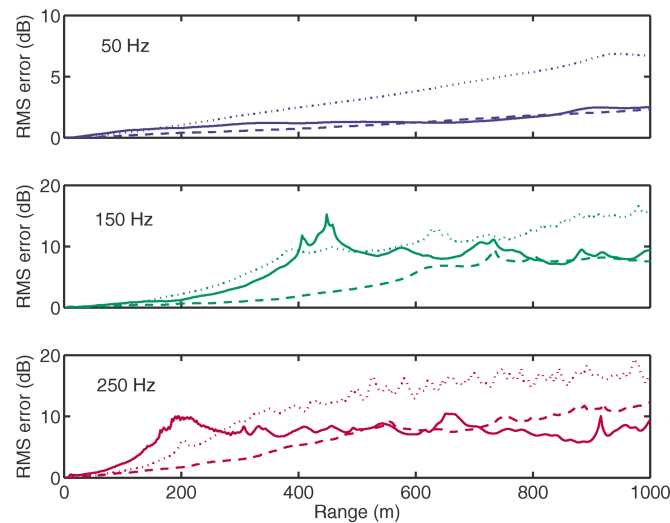
Fig. 4. (Color online) Root mean square errors for estimates of the event sound pressure levels based on instantaneous, displaced vertical profiles. Receiver height is 2 m. Three different frequencies are shown: 50 Hz (top), 150 Hz (middle), 250 Hz (bottom). Solid lines: downwind propagation, stable stratification. Dashed lines: crosswind propagation, neutral stratification. Dotted lines: upwind propagation, unstable stratification.

spheric propagation medium. Of course, other aspects of the outdoor environment not considered here, such as terrain features and ground properties, may also introduce uncertainty into propagation predictions.

Predicting mean sound levels from mean atmospheric profiles results in low errors (typically less than 2 dB), except in shadow and destructive interference zones, where scattering by turbulence is important and sound levels are under predicted. Predictions of mean sound level based on averaging propagation calculations from a large number of vertical profiles samples results in substantial overestimation of levels in shadow zones.

When predicting event propagation, synchronized vertical profiles along the propagation path do *not* lead to good predictive skill, with errors of $8-10$ dB being typical for near-ground propagation. In general, mean atmospheric profiles produce the most accurate predictions of both mean and event sound levels, because they are most representative of typical conditions along the propagation path.

Acknowledgments

References and links

[1] D. K. Wilson, J. M. Noble, and M. A. Coleman, "Sound propagation in the nocturnal boundary layer," J. Atmos. Sci. **60**, 2473–2486 (2003).

[2] P. P. Sullivan, J. C. McWilliams, and C.-H. Moeng, "A subgrid-scale model for large-eddy simulation of planetary boundary-layer flows," Boundary-Layer Meteorol. **71**, 247–276 (1994).

[3] M. West, K. Gilbert, and R. A. Sack, "A tutorial on the parabolic equation (PE) model used for long range sound propagation in the atmosphere," Appl. Acoust. **37**, 31–49 (1992).

[4] P. Blanc Benon, L. Dallois, and D. Juvé, "Long range sound propagation in a turbulent atmosphere within the parabolic approximation," Acust. Acta Acust. **87**, 659–669 (2001).

[5] L. R. Hole and H. M. Mohr, "Modeling of sound propagation in the atmospheric boundary layer: Application of the MIUU mesoscale model," J. Geophys. Res., [Solid Earth] **104**, 11891–11901 (1999).

[6] K. E. Gilbert, R. Raspet, and X. Di, "Calculation of turbulence effects in an upward refracting atmosphere," J. Acoust. Soc. Am. **87**, 2428–2437 (1990).

[7] T. Yokota, K. Makino, Y. Hirao, K. Yamamoto, Y. Okada, and K. Yoshihisa, "Numerical simulation on outdoor sound propagation under the influence of wind speed gradient by the PE method," in *Proceedings of InterNoise 2006*, Honolulu, HI (Institute of Noise Control Engineering of the USA, Inc., 2006).

Fast automatic inharmonicity estimation algorithm

Jukka Rauhala, Heidi-Maria Lehtonen, and Vesa Välimäki
Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, P.O. Box 3000,
FI-02015 TKK, Espoo, Finland.
jukka.rauhala@acoustics.hut.fi, heidi-maria.lehtonen@tkk.fi, vesa.valimaki@tkk.fi

Abstract: A new algorithm is presented for estimating the inharmonicity coefficient of slightly inharmonic stringed instrument sounds. In the proposed partial frequencies deviation method, the inharmonicity is estimated in an intuitive way by minimizing the deviation of the expected partial frequencies compared to the frequencies of the high amplitude peaks in the spectrum. This is done in an iterative process, where the algorithm converges towards the target estimation value. The algorithm is tested using both synthetic and recorded piano tones. The results show that the new algorithm produces accurate results with a small computation cost compared to other methods.
© 2007 Acoustical Society of America

1. Introduction

Inharmonicity is a phenomenon occurring in string instruments due to the stiffness of the string and nonrigid terminations, and it may have an audible effect on the sound produced by the instrument [Järveläinen *et al.*, 2001]. For example, it is generally agreed that the characteristic timbre of the piano is caused in part by the inharmonicity [Fletcher *et al.*, 1962]. The value of the inharmonicity coefficient must be estimated from real data in order to study the auditory perception of the phenomenon or to synthesize these instrument sounds accurately [Bank *et al.*, 2003; Bensa *et al.*, 2005].

The estimation of the inharmonicity value out of recorded data has two main challenges. First, the estimation of the fundamental frequency is required, which can be a difficult task. Second, it is challenging to extract the partial frequencies from real data, because there can be high spectral peaks in the data (due to, e.g., longitudinal modes, or nonlinearities) which do not belong to a single series of partial frequencies. For example, most of the keys of the piano have three strings with slightly different fundamental frequency and inharmonicity values.

A few previous studies have tackled inharmonicity estimation. Galembo and Askenfelt have developed a method based on an inharmonic comb filter realized in the frequency domain to estimate the inharmonicity value [Galembo and Askenfelt, 1999]. Also, cepstral analysis and the harmonic product spectrum, both popular pitch extraction techniques, have been used to estimate inharmonicity [Askenfelt and Galembo, 2000]. Another method for inharmonicity estimation is based on estimating the fundamental frequency in subbands [Klapuri, 2003].

2. Estimation algorithm

Two requirements are defined for the estimation algorithm. First, it should be fully automatic; it should not require parameters other than the fundamental frequency. Second, it should be able to estimate the inharmonicity value from real data with a significant amount of noise, a common problem with recordings of acoustic musical instruments.

Inharmonicity means that the partial frequencies appear to be higher than those of the ideal string. The frequencies of inharmonic partials can be calculated as [Fletcher *et al.*, 1962]
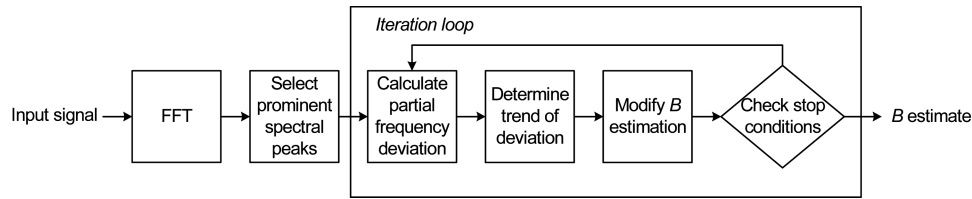
Fig. 1. Structure of the proposed PFD algorithm.

$$f_k = kf_0\sqrt{1 + Bk^2}, \tag{1}$$

where k is the partial number, f_0 is the nominal fundamental frequency of an ideal (nondispersive) string, and B is the inharmonicity coefficient.

Instead of using f_0, which is a theoretical value not existing in the spectrum, the first partial f_1 present in the spectrum is used to calculate the inharmonicity values. Note that when f_0 is substituted with $f_1/(1+B)^{-0.5}$ in Eq. (1), f_0 need not be calculated in order to compute B. This reduces the number of variables to one without affecting the results. A simple technique based on the fast Fourier transform (FFT) is used for estimating f_1, because common pitch detection algorithms, such as YIN [de Cheveigné and Kawahara, 2002], do not work well for inharmonic signals. In this technique, the frequency of the highest amplitude in the spectrum within a given range (see Sec. III) is picked, and the final estimation is done by using three-point parabolic interpolation.

An intuitive way to search for the B value manually would be to plot the spectrum of the signal and the partial frequencies corresponding to the estimated B value. This would be done iteratively, and the B estimate would be modified at each iteration according to the displacement between the partial frequencies and high amplitude peaks in the spectrum to minimize their difference. The algorithm proposed in this letter mimics this method in an automatic way.

Figure 1 shows the block diagram of the proposed algorithm. The first step is to determine the spectrum of the data by calculating the FFT of a long sample from the starting point of the signal. The data are windowed with the Blackman window. The number of FFT points used, 2^{16} points after zero padding in this work, should be large for good resolution. In the second step, the number of the spectral peaks is reduced using a simple technique, where the spectrum is divided into subbands and ten spectral peaks with the highest amplitudes selected from each subband. The width of the subbands is defined to be $5f_1$. Hence, the number of selected peaks is reduced approximately to $2K_{max}$, where K_{max} is the maximum number of partials used in the estimation.

The main part of this algorithm is the iteration loop, which consists of three parts: partial frequencies deviation calculation, determination of the deviation trend, and B estimate (denoted as $\hat{B}$) modification. The initial value of $\hat{B}$ should be within the range of typical B values (10^{-4} to 10^{-3} in piano bass tones), but it does not need to be very accurate as it does not affect much the final result. In this work, the initial value used is $B = 10^{-4}$. In the first part, $\hat{B}$ is used for calculating estimated partial frequencies $\hat{f}_k$ via Eq. (1). In order to avoid bias due to the limited frequency resolution, $\hat{f}_k$ are quantized using the frequency resolution due to FFT. Then, the frequencies f_k corresponding to the highest amplitudes within the closed interval $[\hat{f}_k - \Delta f, \hat{f}_k + \Delta f]$ are selected, and the partial frequency deviation $D_k = \hat{f}_k - f_k$ is calculated. In this work, the parameter $\Delta f = 0.4f_1$.

The next task in the iteration loop is to determine the trend of the deviation which indicates whether $\hat{B}$ should be decreased or increased. The problem is that the deviation curve is not always smooth and there can be sharp peaks and notches in the curve. This is taken into account by defining the trend to be positive, if there are more positive derivates than negative
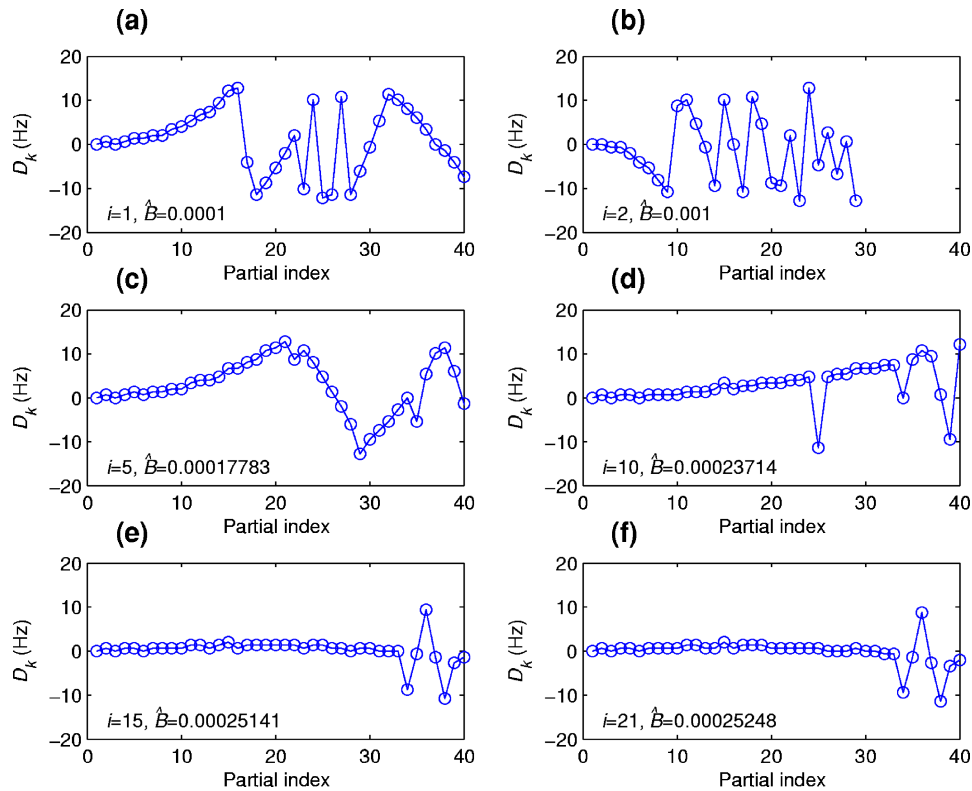
Fig. 2. (Color online) The partial frequency deviation D_k curve at iteration loop rounds (a) 1, (b) 2, (c) 5, (d) 10, (e) 15, and (f) 21, when a recorded Bb_0 tone is analyzed ($f_0 = 32.4$ Hz). The corresponding $\hat{B}$ values at each round are shown in the figures, the final B estimate is 0.0002546.

derivates, and vice versa. Hence, the trend can be calculated, for example, as $\text{sign}(\Sigma_{k=2}^{K_{\max}} \text{sign}(D_{k+1} - D_k))$. The latter part of the iteration is the modification of $\hat{B}$, based on the determined trend of the deviation. The $\hat{B}$ value is changed by multiplying it by 10^δ, where δ is an adaptive step-size parameter (the initial value is set to 1 in this work), which has a positive sign if the trend is positive, or a negative sign if it is negative. Whenever a reverse occurs (that is, two adjacent trends have different signs), parameter δ is modified by dividing it by 2. In other words, the goal is to increase (or decrease, depending on the sign of the determined trend) $\hat{B}$ until it is larger than the target B estimate, which can be seen in the D_k curve as a change of trend. Then, the iteration loop is performed towards the opposite direction with a decreased step size, assuming that the iteration loop converges around the target B estimate value. Figure 2 and Mm. 1 illustrate how the deviation curve behaves in different stages of the iteration loop. The trade-off between the speed of the algorithm and the accuracy of the algorithm can be controlled by the loop stop conditions. This work uses the following stop conditions: the iteration loop is run until either it has been run for 40 times or $|\delta| < 10^{-4}$. Finally, the proposed method provides a way to evaluate the produced B estimate reliably, as the smoothness of the last D_k curve indicates the level of accuracy.

Mm. 1. Algorithm convergence with a synthetic signal (1.3 MB). This is a file of type "mpeg."

In addition to B estimation, the D_k curve provides information on the accuracy of the f_1 estimate: convexity in the curve indicates that the estimate should be increased and concavity indicates the need to decrease the estimate. This can be used for refining the f_1 estimate, which

improves the accuracy of B estimation as well. The B estimation by using the f_1 estimate refining is done as follows. First, the B estimation is done as proposed above to obtain the initial B estimate value. Then, the same algorithm, as shown in Fig. 1, is used for f_1 estimate refining, only this time instead of analyzing the trend of the deviation and modifying B, the convexity and concavity of the deviation is estimated, and the f_1 estimate is modified. The convexity and concavity can be estimated by examining the average of the first half of D_k values: the positive curve is convex and vice versa. Based on this, the f_1 estimate is modified by multiplying it with $(1+\mu)$, where μ is the step size, which has a positive sign if the curve is convex and a negative sign if the curve is concave. The initial value of μ is set to 0.005. Similarly to δ, μ is divided by 2 at each change in trend. The refining iteration is stopped when it has been run for 100 times or $|\mu| < 10^{-5}$. After the f_1 estimate is refined, the original B estimation algorithm is run again to modify the B estimate.

3. Results and comparison

The proposed algorithm was compared against two versions of the inharmonic comb filter (ICF) method [Galembo and Askenfelt, 1999]. In the ICF method, which was included in the comparison because it is a good previously introduced method, the signal is filtered with a multi-band filter. The center frequencies of the pass bands are determined from Eq. (1). The ICF method does not need an external f_0 estimation method, as it combines the inharmonicity estimation and the f_0 estimation.

In the first phase of the ICF method, a rough estimation of f_0 is made with a pre-defined value for the inharmonicity. Then, the rough f_0 estimate is used to calculate a rough estimation for the inharmonicity value. Both f_0 and B values are varied and a power spectrum sum is calculated at each point. The estimated inharmonicity value is determined to be the value corresponding to the maximum power spectrum sum. In the final phase, the ranges of the B and f_0 values are narrowed down around the rough estimates obtained in the previous phases. At the end, the result from the third phase is refined by parabolic interpolation. The details of the algorithm are documented in [Galembo and Askenfelt, 1999].

In order to improve the performance of the ICF method in the tests, the parameters suggested in [Galembo and Askenfelt, 1999] were slightly modified (the modified version is denoted in this work as ICF+). First, the bandwidth of the pass band was chosen to be 5% of f_0. Second, a linear scale was used in the second phase of the algorithm to produce better results than with the logarithmic scale proposed in [Galembo and Askenfelt, 1999]. Moreover, the number of the B values on the grid in the second phase was changed to 400, and in the third phase it was chosen to be 100 according to test in this work.

The number of partials included in the estimation was chosen to be 50 for the proposed algorithm, whereas for the ICF and ICF+ it was 30, as suggested in [Galembo and Askenfelt, 1999]. The sampling rate used was 44,100 Hz in all cases. Synthetic signals generated with additive synthesis were used in the first case. The correct fundamental frequency used in synthesis was given to all methods in order to avoid the bias caused by the fundamental frequency extraction. The second case tested the inharmonicity estimation with real piano data. In this case, each method used its own fundamental frequency estimation to determine the fundamental frequency from the data (PFD was run with the f_1 estimate refining, as described in Sec. II). Since the piano key number is known in all cases, the methods assumed that the correct f_0 lies within the range a major third up and down from f_0 (based on an equal-tempered scale with $A_4 = 440$ Hz) in order to prevent octave errors. The test cases were limited to key range 1–35, since the original ICF method is designed for $f_0 < 200$ Hz. Moreover, the inharmonicity effect is most important in the bass range of the piano [Galembo and Askenfelt, 1999].

The first test case simulated the inharmonicity estimation of the piano sounds. A synthetic signal was created for each key of the piano. A typical inharmonicity value of the piano was selected for each signal, see, e.g., [Askenfelt and Galembo, 2000]. The partial amplitudes were measured from real data, with some manual fine tuning, and a variation of 0–10% was added to the synthetic partial amplitudes. Each key included the correct number of synthesized
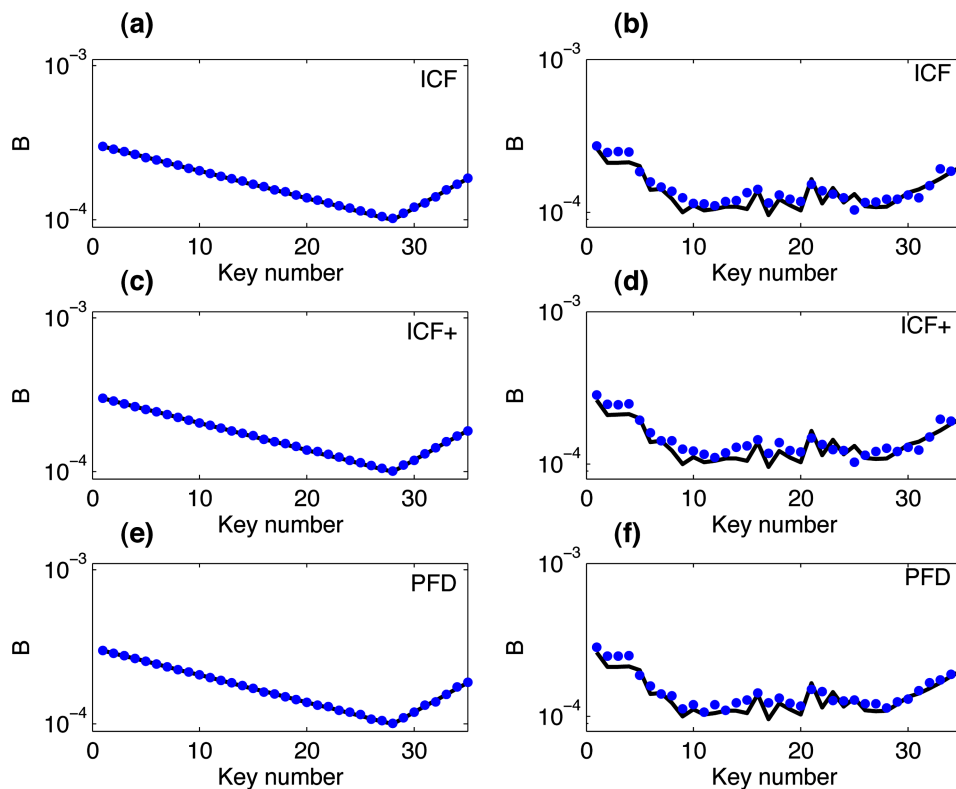
Fig. 3. (Color online) Inharmonicity estimation test for synthetic piano data with SNR of 40 dB using (a) ICF, (c) ICF+, and (e) PFD. The solid line represents the correct inharmonicity coefficient values for keys 1-35. The same test was done with recorded piano tones[1] by using (b) ICF, (d) ICF+, and (f) PFD. The solid line indicates manually estimated values.

strings, i.e., keys 1–12 had only one string, whereas keys 13–29 had two, and keys 30–35 had three strings. The signal-to-noise ratio (SNR) was determined by comparing the power of the signal during its first 10 ms to the power of the noise in a 10 ms window. Gaussian white noise was added into the signals so that a SNR of 40 dB was obtained during the first 10 ms of the signal. The test signals are available at http://www.acoustics.hut.fi/publications/papers/pfd. Finally, the three methods were tested by using real piano samples.[1] The SNR in these piano tones was approximately 35 dB.

        Figure 3 shows the test results for the two test cases. In addition, the total running times and average rms error values corresponding to the first test case and average rms deviation values for the test case using recorded tones are shown in Table 1. The results from the first test case, which used synthetic piano tones, show that all methods produce good results. The differences are very small and the methods can be considered equal in terms of the rms error. In the second test case, PFD had the best performance according to the rms values. Moreover, ICF and ICF+ produced slightly less accurate results. The differences in the computational cost between the presented method and the inharmonic comb filter method variations were remarkable: the running time of the test case using the real piano tones for ICF+ was over 15 min, whereas PFD produced better results in less than 1 min. Hence, the proposed method was proven to produce good results in a fraction of the time compared to other methods.

4. Conclusion

An algorithm for estimating the inharmonicity coefficient for the sounds of stringed musical instruments was presented. The algorithm is simple and it is able to produce fairly accurate

Table 1. Total running times of test cases, and average RMS errors (synthetic tones) and average RMS deviations (real tones) of the estimated inharmonicity values for the three methods.

Method	Synthetic tones		Real tones	
	RMS error	Running time	RMS deviation[a]	Running time
ICF	1.19×10^{-6}	11.4 s	1.73×10^{-5}	313.9 s
ICF+	1.16×10^{-6}	24.1 s	1.88×10^{-5}	946.5 s
PFD	1.19×10^{-6}	7.0 s	1.66×10^{-5}	58.4 s

[a]Difference from manually estimated values. The correct inharmonicity value is unknown.

estimations at a small computational cost. In addition, it is able to refine fundamental frequency estimation. The proposed algorithm can be used as an analysis tool to obtain a good estimate of the inharmonicity of various string instruments. Future work includes examination of improving the accuracy of the B estimation, and study on using the algorithm for f_0 estimation with stringed instrument tones.

Acknowledgment

References and links

[1]Steinway grand piano samples from University of Iowa (http://theremin.music.uiowa.edu Viewed 4/2/07).

Askenfelt, A., and Galembo, A. S. (**2000**). "Study of the spectral inharmonicity of musical sound," Acoust. Phys. **46**(2), 121–132.

Bank, B., Avanzini, F., Borin, G., De Poli, G., Fontana, F., and Rocchesso, D. (**2003**). "Physically informed signal processing methods for piano sound synthesis: A research overview," EURASIP J. Appl. Signal Processing **2003**(10), 941–952.

Bensa, J., Bilbao, S., Kronland-Martinet, R., Smith III, J. O., and Voinier, T. (**2005**). "Computational modeling of stiff piano strings using digital waveguides and finite differences," Acta. Acust. Acust. **91**(2), 289–298.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**(4), 1917–1930.

Fletcher, H., Blackham, E. D., and Stratton, R. (**1962**). "Quality of piano tones," J. Acoust. Soc. Am. **34**(6), 749–761.

Galembo, A. S., and Askenfelt, A. (**1999**). "Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano," IEEE Trans. Speech Audio Process. **7**(2), 197–203.

Järveläinen, H., Välimäki, V., and Karjalainen, M. (**2001**). "Audibility of the timbral effects of inharmonicity in stringed instrument tones," Acoust. Res. Letters Online, **2**(3), 79–84.

Klapuri, A. (**2003**). "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," IEEE Trans. Speech Audio Process. **11**(6), 184–194.

Simulating the acquisition of lexical tones from continuous dynamic input

Bruno Gauthier and Rushen Shi

Département de psychologie, Université du Québec à Montréal, C.P. 8888, Succursale Centre-Ville, Montréal, Québec H3C 3P8, Canada
gauthier.bruno@courrier.uqam.ca, shi.rushen@uqam.ca

Yi Xu

Department of Phonetics and Linguistics, University College London, Room G32, Wolfson House, 4 Stephenson Way, London, NW1 2HE, United Kingdom
yi@phon.ucl.ac.uk

Abstract: Infants develop phonetic categories by simply being exposed to adult speech. It remains unclear, however, how they handle the extensive variability inherent to speech, and how they process multiple linguistic functions that share the same acoustic parameters. Across four neural network simulations of lexical tone acquisition, self-organizing maps were trained with continuous speech input of increasing variability. Robust tonal categorization was achieved by tracking the velocity profiles of fundamental frequency contours. This result suggests that continuous speech signal carries sufficient categorical information that can be directly processed, and that dynamic acoustic information can be used for resolving the variability problem.

1. Introduction

Before reaching 1 year of age, infants have developed the ability to process speech sounds specific to their native language. Infants are born with general auditory mechanisms to process all speech sounds of human languages,[1,2] and later show a decline in sensitivity to non-native categories while narrowing in on the native language phonetic categories.[3–5] This development seems to be related to their sensitivity to the statistical distribution of speech sounds, as it was recently demonstrated that 6–8-month-old infants develop discrimination sensitivity corresponding to bi-modal distribution of VOT values in the stimuli after brief training.[6] However, the speech input to infants is much more variable than bi-modal distributions. Speech signal is continuous and dynamic, and steady-state patterns are rare.[7] There is extensive contextual variability due to coarticulation[8] and large amount of cross-speaker variations. Moreover, there are often more than two phonetic categories along any particular acoustic/articulatory dimension, and infants do not know in advance how many of them there are in a given language. It thus remains unclear how the early perceptual system can develop phonetic categories from such highly variable acoustic input.

Here we study with unsupervised neural networks whether phonetic categories can be discovered by directly processing continuous speech signals that contain different types of variability and competing linguistic functions. We simulate the acquisition of lexical tones in Mandarin. Mandarin has four tones for distinguishing words that can be identical in segmental compositions: High (H), Rise (R), Low (L), and Fall (F), which are carried by the fundamental frequency (F_0) of the vocal fold vibrations. Because tones typically involve a single primary acoustic dimension, namely, F_0, they are ideal for testing hypotheses about detailed mechanisms of phonetic acquisition.

The F_0 values of a tone, however, vary extensively due to at least three sources (Fig. 1). First, cross-speaker variability arises from differences such as age, gender, and idiosyncrasies (e.g., Refs. 9 and 10). Second, contextual variability arises from neighboring sounds affecting
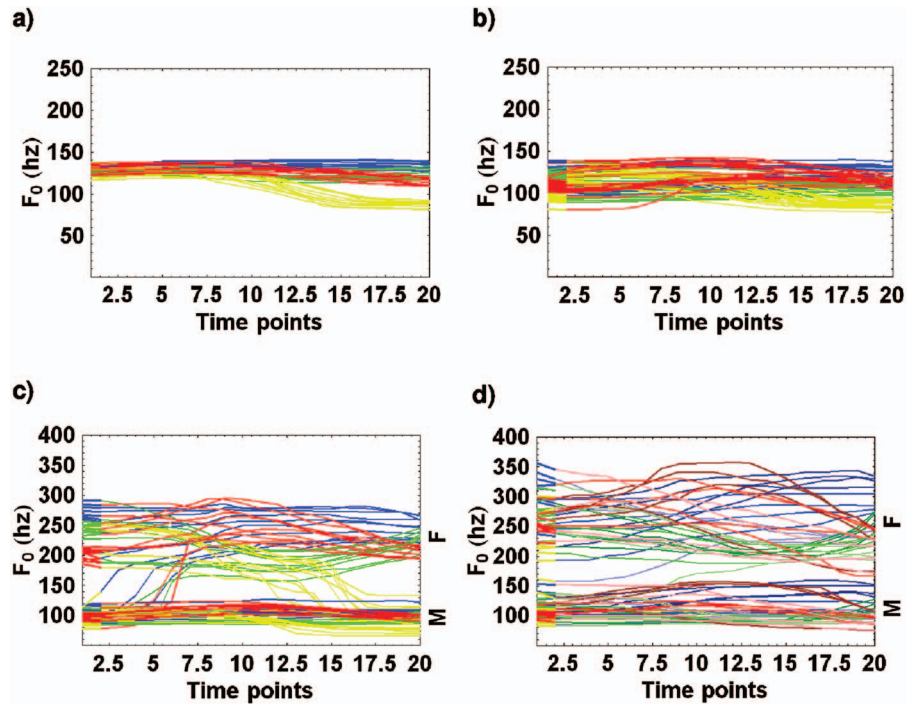
Fig. 1. Variability in tones. (a) F_0 (in hertz) of 40 repetitions of the four Mandarin tones (High, Rise, Low, Fall) by one male speaker with identical preceding tone (High) and identical focal status (neutral). (b) Contextual variability: F_0 of 80 repetitions of the four tones by one male speaker with different preceding tones (specified by syllable onset color) and identical focal status (neutral). (c) Contextual and speaker variability: F_0 of 80 repetitions of the four tones by one female and one male speaker with different preceding tones and identical focal status. (d) Contextual, speaker, and focal variability: F_0 of 80 repetitions of the four tones by one female and one male speaker with different preceding tones and variable focus (dark=on-focus, medium=neutral focus, pale=post-focus).

one another.[11] The F_0 pattern of any tone varies extensively due to the mechanical carryover influence of the preceding tone.[12] For example, the F_0 contour of a High tone following a Low tone somewhat resembles that of a Rise tone. Speech addressed to infants consists primarily of multiword utterances,[13] leading to considerable contextual variability. Finally, variability in F_0 comes from its use to not only distinguish words, but also to encode information such as focus (for emphasizing part of a message), which can introduce F_0 variations with magnitudes similar to those of tones.[14]

Despite the extensive variability, tonal perception appears early in infancy.[15] A key to infants' strategies may lie in the understanding of the mechanism of contextual variability; that is, despite the apparent variability, the talkers' articulatory strategy remains the same: to approach a constant tonal target starting from the syllable-initial F_0 left by the tone of the preceding syllable.[16] Such a strategy would result in velocity profiles that directly reflect the nature of the tonal targets. Velocity profiles have been shown to reveal the dynamics of skilled actions such as jaw movements during speech.[17] Moreover, taking the derivative of a curve leads to the removal of all its constant term(s), eliminating any overall height differences such as those due to cross-speaker variability. Thus, our hypotheses are (a) tonal categories can be discovered by processing syllable-sized continuous pitch movement patterns and (b) the velocity of F_0 (i.e., the first derivatives of F_0 patterns, henceforth D_1), better reveals the invariant properties of tones than F_0.

2. Method

In four simulations, self-organizing-maps [SOMs (Ref. [18])] were presented with learning material of increasing degrees of variability to assess their impact on tonal categorization. Self-organizing networks have been useful for characterizing the mechanism underlying various language acquisition tasks (e.g., Refs. 13 and 19) and for solving statistical pattern recognition problems. Given infants' sensitivity to distributional properties of speech signals,[6] the SOM is ideal for modeling the perceptual learning of phonetic categories. The SOM combines a map of topologically ordered processing units with a high-dimensional input space. Each output unit is connected to the input space by an adaptive weight vector, the dimensionality of which corresponds to that of input vectors. When an input token is presented to the network, the winning unit (i.e., with the shortest Euclidean distance to the input vector) is activated, and its connection weights are moved toward the data point by the learning rule. After training, the SOM is expected to reveal the structure of the data.

In the present study, the map units were arranged in a square topology, and initial connection weight values were arbitrarily assigned to cover only a small portion of the input space. The input space was formed of a set of continuous F_0 or D_1 vectors. In simulation 1, the input corpus contained 1800 exemplars produced by three adult male native Mandarin speakers (data from Ref. 12). Each stimulus corresponded to the first or second syllable of the word "mama" produced with varying tones in the middle of a carrier sentence that had either high or low pretarget F_0 offset and posttarget F_0 onset. In simulations 2, 3, and 4, tonal exemplars were from 3840 declarative sentences produced by eight native Mandarin speakers (data from Ref. 14). Sentences were formed of a subject, verb, and object and contained five syllables, each word corresponding to one or two consonant-vowel (CV) syllable(s), where C was a sonorant (/m,n/), except when the Low tone occurred on the fourth syllable, where C was /d/. The subject and object words were disyllabic and the verb was monosyllabic. The sentences were produced in various focus conditions: (a) neutral focus, (b) focus on word 1, (c) focus on word 2, and (d) focus on word 3 [e.g., "maomi mo maomi" (kitty touches kitty)] in response to the following wh-questions: "What is Kitty doing?" "Who is stroking Kitty?," "What is Kitty doing to Kitty?," "What is Kitty stroking?"). Since the tone on the first and last syllables was kept constant to High, these syllables were removed from the input corpus. The second, third, and fourth syllables contained varying tones (H, R, L, F on the second syllable, H, R, F on the third syllable, and H, L on the fourth syllable). Simulations 2–4 thus involved stimuli produced in all possible tonal contexts and focus conditions. Simulation 2 involved 1440 exemplars produced by a single speaker, and simulation 3 used 5760 tones produced by four male speakers. Simulation 4 involved the highest amount of variability, with 11 520 tones produced by four male and four female speakers in all tonal and focus conditions.

Input tokens to the network were 20-point vectors composed of equal-distanced discrete values from syllable-sized F_0 curves. The continuous F_0 contour was extracted by taking the inverse of every vocal period (for the detailed F_0 extraction procedure, see Ref. 14). F_0 input vectors were first transformed from hertz scale to semitone scale according to

$$F_0 st = 12 \log_2 (F_0 hz) \tag{1}$$

The velocity profiles of F_0 were generated according to

$$D_{1i} = (F_0 st_{i+1} - F_0 st_i)/(T_{i+1} - T_i) \tag{2}$$

where T represents time, which yields the discrete first derivatives of F_0. Each simulation presented one network with F_0 and one network with D_1 to compare the performance of both parameters. D_1 was expected to yield better results than F_0 in all simulations, and focused syllables were expected to be categorized better than postfocused syllables.

To compare the results obtained in different simulations, the size of a map was proportional to the input corpus size by a factor of about $\frac{1}{10}$ [e.g., 144 (12×12) units for Simulation 2 (1440 tokens)]. During training, half of the input tokens were randomly presented to the networks as whole vectors. The learning step size decreased linearly from 0.7 to 0.01. The neigh-
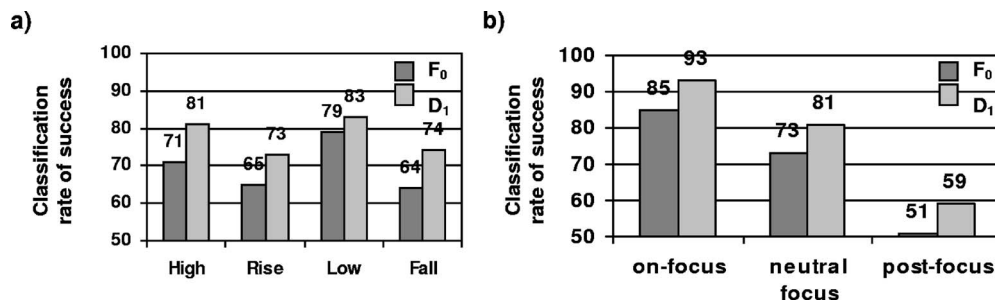
Fig. 2. Percent correct classification of F_0 (dark bars) and D_1 (pale bars) networks during simulation 4. (a) Results for tones High, Rise, Low, and Fall. (b) Results expressed as a function of focal status (on-focus, neutral focus, and post-focus).

borhood activation function included all units at the beginning, decreased exponentially, and only included the winning unit toward the end of training. The testing phase presented the training corpus and new input tokens to verify the networks' capacity to generalize to novel data. During testing, units that responded to a single category at least 68% of the time were labeled as that category. Units that responded to multiple tonal categories, none of which was dominant, were treated as noncategorical. Learning was assessed in terms of classification rate of success, i.e., the percentage of correctly classified input tokens.

To better understand the outcome of the categorization process, quantitative coloring of the trained maps was obtained by associating each tonal category with a distinct color produced with the CMYK color system (the High tone is represented by blue, i.e., a mix of cyan and magenta in the vector [1,1,0,0]; Rise=green [1,0,1,0]; Low=yellow [0,0,1,0]; Fall=red [0,1,1,0]). Each map unit was then associated to a four-dimensional vector, the values of which were specified according to the unit firing probabilities for each tonal class during testing [the fourth element K (black) was kept null]. As a consequence, units responding to a single tone are represented by a saturated color while noncategorical units are represented by "impure" colors. Learned categories are thus shown on the maps as regions of distinct colors.

3. Results

Overall, the networks trained with either F_0 or D_1 yielded above chance level performance, but those trained with D_1 performed better than those trained with F_0. In simulation 1 (input from different tonal contexts by three male speakers) a reasonably high rate of success was achieved for each tone with F_0 (H: 73%, R: 84%, L: 96%, F: 83%; mean: 84%, standard deviation: 9%), but D_1 yielded almost perfect categorization (H: 93%, R: 96%, L: 94%, F: 90%; mean: 93%, s.d.: 3%), replicating our previous study.[20] In simulation 2 (input from one speaker in different tonal contexts with variable focus, i.e., on-focus, neutral focus, and postfocus) the overall rate of success decreased relative to simulation 1 for F_0 (H: 68%,R: 69%, L: 88%, F: 63%; mean: 72%, s.d.: 11%) and for D_1 (H: 85%, R: 81%, L: 87%, F: 84%; mean: 84%, s.d.: 3%), suggesting that focus-induced variability is more detrimental to tonal categorization than cross-speaker (within-gender) variability. Simulation 3 tested the combined impact of the aforementioned sources of variability (input was from four male speakers in different tonal contexts with variable focus). Although the performance of both networks declined, D_1 (H: 82%, R: 76%, L: 77%, F: 79%; mean: 79%, s.d.: 3%) still performed better than F_0 (H: 64%, R: 59%, L: 75%, F: 70%; mean: 67%, s.d.: 7%). Finally, the results of simulation 4 (cross-gender, contextual, and focus induced variability) are shown in Fig. 2. Both F_0 (mean: 70%, s.d.: 7%) and D_1 (mean: 78%, s.d.: 5%) declined in performance relative to previous simulations, with D_1 still showing superiority. Most of the errors in simulation 4 involved postfocused elements [Fig. 2(b)]. The rate of success of the D_1 network for on-focus syllables remained excellent (93%), consistent with the expectation that focused elements should be perceptually salient.
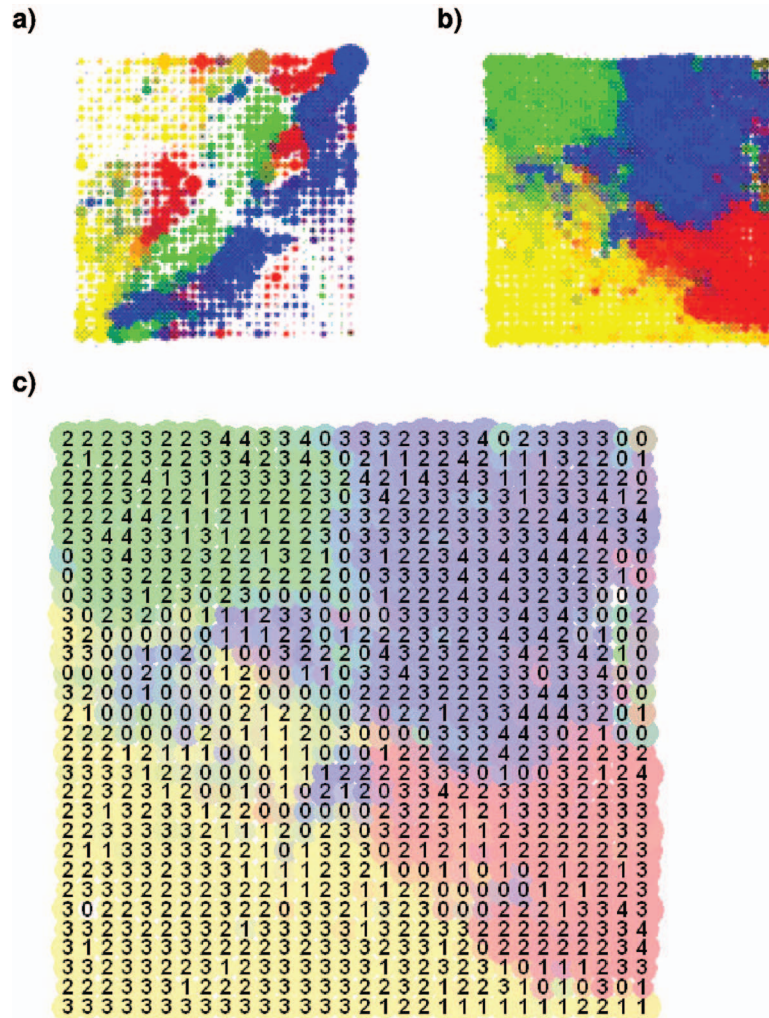
Fig. 3. Color maps of F_0 (a) and D_1 (b) networks in simulation 4 (High, Rise, Low, Fall). (c) Tonal context map for D_1 network, where categorical units are represented by the number of tonal contexts they are sensitive to (1 to 4; zeros correspond to noncategorical units).

Tonal color maps of simulation 4 are shown in Fig. 3. The F_0 color map [Fig. 3(a)] exhibits no clear tonal organization. Some tonal categories are distributed to multiple clusters. Activations of units during testing are uneven across the map: some units responded to many input tokens, while others to few or none (the larger a unit, the greater its firing probability). In contrast, the D_1 color map [Fig. 3(b)] shows four well-separated regions—each corresponding to one tone. Activations of units are even: each responding to a comparable number of test tokens. Verification of the response patterns of units shows that most categorical units responded to tones in multiple preceding contexts [Fig. 3(c)], confirming that each region indeed forms a single category and does not break down into subclusters based on tonal contexts (e.g., no subcluster for High tones preceded by other High tones). D_1 is thus much more powerful than F_0 in normalizing and categorizing the Mandarin tones.

4. Discussion and conclusion

The results of our simulations show that, despite extensive variability due to context, speaker, and competing linguistic functions (here tones and focus), a simple inductive learning mecha-

nism can extract tonal categories directly from continuous acoustic input without any supervision or feedback, assuming that F_0 contours have already been segmented into syllable-sized chunks by a separate mechanism.[21] We also find that velocity profiles as input for phonetic categorization are more robust than F_0 contours in handling the variability and revealing the invariant underlying phonetic targets. The effectiveness of using continuous dynamic patterns as input also eliminates the need for pre-extracting any summary properties as categorical cues, as such cues, if any, seem to operate implicitly rather than explicitly in the simulated learning process. Granted, no one has yet shown that infants, or even adults, have the ability to compute velocity profiles from acoustic signal. What the results of our simulations have demonstrated is that, if they did, the benefit would be enormous. It would thus be desirable for future research to look for direct evidence of neural processing of velocity.

Acknowledgments

References and links

[1] R. N. Aslin, J. F. Werker, and J. L. Morgan, "Innate phonetic boundaries revisited," J. Acoust. Soc. Am. **112**(4), 1257–1260 (2002).

[2] P. D. Eimas, J. L. Miller, and P. W. Jusczyk, "On infant speech perception and the acquisition of language," in *Categorical Perception: The Groundwork of Cognition*, edited by S. Harnad (Cambridge University Press, New York, 1987), pp. 161–195.

[3] J. F. Werker and R. C. Tees, "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**(1), 49–63 (1984).

[4] P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom, "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**(5044), 606–608 (1992).

[5] P. K. Kuhl, E. Stevens, A. Hayashi, T. Deguchi, S. Kiritani, and P. Iverson, "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months," Developmental Sci. **9**(2), F13–F21 (2006).

[6] J. Maye, J. F. Werker, and L. Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**(3), B101–B111 (2002).

[7] J. S. Perkell and D. H. Klatt, *Invariance and Variability in Speech Processes* (Erlbaum, Hillsdale, NJ, 1986).

[8] W. J. Hardcastle and N. Hewlett, *Coarticulation: Theory, Data and Techniques* (Cambridge University Press, Cambridge, 1999).

[9] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**(2), 175–184 (1952).

[10] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111 (1995).

[11] S. E. G. Ohman, "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. **39**(1), 151–168 (1966).

[12] Y. Xu, "Contextual tonal variations in Mandarin," J. Phonetics **25**, 61–83 (1997).

[13] R. Shi, J. L. Morgan, and P. Allopenna, "Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective," J. Child Lang **25**(1), 169–201 (1998).

[14] Y. Xu, "Effects of tone and focus on the formation and alignment of F_0 contours," J. Phonetics **27**, 55–105 (1999).

[15] K. J. Mattock, "Perceptual reorganisation for tone: Linguistic tone and non-linguistic pitch perception by English language and Chinese language infant," Unpublished doctoral dissertation, University of Western Sydney, 2004.

[16] Y. Xu and Q. E. Wang, "Pitch targets and their realization: Evidence from Mandarin Chinese," Speech Commun. **33**, 319–337 (2001).

[17] W. L. Nelson, "Physical principles for economies of skilled movements," Biol. Cybern. **46**(2), 135–147 (1983).

[18] T. Kohonen, *Self-Organizing-Maps* (Springer, Berlin, 1995).

[19] F. H. Guenther and M. N. Gjaja, "The perceptual magnet effect as an emergent property of neural map formation," J. Acoust. Soc. Am. **100**(2), 1111–1121 (1996).

[20] B. Gauthier, R. Shi, and Y. Xu, "Learning phonetic categories by tracking movements," Cognition **103**(1), 80–106 (2007).

[21] J. Bertoncini and J. Mehler, "Syllables as units in infant speech perception," Infant Behav. Dev. **4**(3), 247–260 (1981).

Effects of auditory feedback deprivation length on the vowel /ɛ/ produced by pediatric cochlear-implant users

Sneha V. Bharadwaj, Amanda G. Graves, Delia D. Bauer, and Peter F. Assmann
School of Behavioral and Brain Sciences, Callier Center for Communication Disorders,
University of Texas at Dallas, 811 Synergy Park Blvd, Richardson, Texas 75080
snehab@utdallas.edu, agg043000@utdallas.edu, dda051000@utdallas.edu, assmann@utdallas.edu

Abstract: Effects of auditory deprivation on speech production by ten cochlear-implanted children were investigated by turning off the implant for durations ranging from 0.3 to 5.0 s and measuring the formant frequencies ($F1$ and $F2$) of the vowel /ɛ/. In five of the ten talkers, $F1$ and/or $F2$ shifted when auditory feedback was eliminated. Without feedback, $F2$ frequency lowered consistently, suggesting vowel centralization. Phonetic transcription indicated that some of these acoustic changes led to perceptible shifts in phonetic quality. The results provide evidence that brief periods of auditory deprivation can produce perceptible changes in vowels produced by some cochlear-implanted children.
© 2007 Acoustical Society of America

1. Introduction

Auditory feedback is considered important for speech development and maintenance (e.g., Borden, 1979; Perkell *et al.*, 1992). While several studies of deaf and normal-hearing individuals have attributed a calibrational role for auditory feedback in speech production within the short term (e.g., Zimmermann and Rettaliata, 1981), the exact nature of its role continues to be a topic of investigation. Studies of deaf individuals have employed the processor-on versus -off paradigm (e.g., Svirsky and Tobey, 1991; Svirsky *et al.*, 1992) as it allows an investigator to examine various aspects of speech produced in the presence and the complete absence of auditory feedback. This approach also permits the researcher to control the length of time for which auditory feedback is present or absent. Changes in both suprasegmental and segmental aspects of speech have been documented when auditory deprivation varied from 20 s–24 h (e.g., Svirsky and Tobey, 1991; Matthies *et al.*, 1996). Moreover, a few studies have demonstrated that auditory feedback-related changes in speech (e.g., fricative consonants, voice quality) produced by some talkers are perceptible to normal-hearing listeners (Tartter *et al.*, 1989; Bharadwaj *et al.*, 2006). While studies of acquired hearing impairments have shown gradual deterioration in speech quality consequent to hearing loss, studies using on-off paradigms or off-on-off paradigms have noted rapid changes in speech following auditory deprivation and restoration (see Svirsky *et al.*, 1992). These rapid changes suggest that the speech production mechanism needs to be updated with feedback information in order to fine-tune its settings and that the absence of feedback leads to drifting in speech parameters (e.g., Matthies *et al.*, 1996). In addition, these rapid changes in speech may represent either a deliberate strategy used by talkers who attempt to produce clear speech in the absence of auditory feedback or a switch from the use of the current internal model to the one acquired before cochlear implantation (see Bharadwaj *et al.*, 2006, for a summary).

Another approach to investigating the short-term effects of auditory feedback is to examine the time window involved in adjusting/correcting parameters of speech in response to altered auditory feedback in listeners with normal hearing (e.g., Kawahara and Williams, 1996;

Houde and Jordan, 1998; Jones and Munhall, 2000; Xu *et al.*, 2004; Purcell and Munhall, 2006b). Studies of normal-hearing individuals have shown adjustments in fundamental frequency ($F0$) over the length of a vowel or syllable, approximately 150 ms following manipulations to auditory feedback. While some investigators suggest that the time window of 150 ms would be too brief to assist in the online control of $F0$ (e.g., Donath *et al.*, 2002), other researchers propose that auditory feedback is used to regulate $F0$ across the length of a syllable (e.g., Xu *et al.*, 2004). In most of the auditory feedback perturbation studies, manipulation to auditory feedback has primarily involved upward or downward shifts in $F0$ only. There are a limited number of studies investigating the effects of online perturbation of auditory feedback on other speech attributes such as the formant frequencies, $F1$ and $F2$. An exception is Purcell and Munhall (2006a) who showed compensations in the production of the vowel /ɛ/ by normal-hearing adults in response to an unexpected increase or decrease in $F1$. The compensation was estimated to begin less than 460 ms following manipulation to auditory feedback, suggesting that the time involved in implementing auditory feedback-based corrective changes to $F1$ is longer than for $F0$. These findings raise important questions about the minimal time needed to make auditory feedback-based corrective changes in various types of speech parameters. Such data from children are quite limited and might make a valuable contribution to our understanding of the role of auditory feedback in the development of internal models for speech (see Perkell *et al.*, 2000; Guenther, 2006).

　　　The present study examined whether brief deprivation of auditory feedback ranging from $0.3-5.0$ s leads to significant shifts in $F1$ and $F2$ of the vowel /ɛ/ produced by children who were fitted with cochlear implants. An additional objective of this study was to investigate whether the acoustic effects that occur in vowels in the absence of auditory feedback are perceptible to trained listeners.

2. Acoustic analyses and phonetic transcription

2.1 Methods

2.1.1 Participants

Participants were ten prelingually deafened children who were fitted with multichannel cochlear implants. Information concerning age, gender, age of implantation, length of implant use, type of implant, implanted ear, and speech intelligibility is reported in Table 1. All participants had severe-profound hearing loss in the unimplanted ear. Participants were monolingual speakers of American English and used the oral-aural mode of communication. Participants were paid for their participation. The range of speech intelligibility scores (production) for each talker is reported in Table 1 and was assessed using the procedure described in Tobey *et al.* (2003). In this procedure, a set of three different listeners heard 36 sentences spoken by each talker and wrote down what they heard, guessing if necessary. The scoring was based on the total number of words in a sentence understood correctly by listeners.

2.1.2 Speech materials and procedures

The speech materials for this experiment included three sentences each with increasing phonetic material preceding the target vowel: (1) "a **head** again," (2) "This is a nose and a **head** again," and (3) "Hey look here, this is a nose and a **head** again." The target vowel was /ɛ/ in the word "head." This vowel was selected because several studies have shown significant formant frequency shifts when it is produced in the absence of auditory feedback (e.g., Svirsky and Tobey, 1991). Across the three sentence types, vowel /ɛ/ is increasingly distant (approximately $0.3-5.0$ s) from the sentence onset.

　　　Participants read a randomized list of 12 repetitions of each sentence in two conditions: cochlear implant processor-on (auditory feedback present) and processor-off (auditory feedback absent). The processor was turned on or off manually by the experimenter at the beginning of each sentence in a predetermined random order. Participants did not speak anything other than the target sentences. The sentences were audio recorded in a sound-treated room

Table 1. Demographic information for early-implanted and late-implanted participants.

Talker	Age (yrs)	Sex	Implanted age (yrs)	Length of CI use (yrs)	CI type and/ implanted ear	Range of speech intelligibility score (%)	Speech perception score (H.I.N.T.)
T1	7:2	F	2:1	5:1	Advanced bionics/L	95–96	89
T2	10:0	F	2:1	7:11	Nucleus 24/ L,R	99–100	86
T3	15:3	M	3:7	11:5	Nucleus 22/L	89–91	68[a]
T4	7:9	F	3:9	4:0	Advanced bionics/R	90–94	97
T5	10:4	M	2:7	7:9	Advanced bionics/R	81–91	56
T6	12:0	F	5:11	6:1	Nucleus 24/ L	95–98	98
T7	13:10	F	5:0	8:10	Nucleus 22/L	83–96	94
T8	10:6	F	5:1	5:5	Med-El C40+/L	94–96	88
T9	12:4	M	9:9	2:6	Nucleus 24/R	56–72	74
T10	9:7	M	5:4	4:3	Nucleus 22/L	83–96	88[b]

[a]PBK-50.
[b]Lexical Neighborhood Test.

using Sony PCM-M1 Digital Audio Recorder and a Sony ECM-719 condenser microphone placed approximately 10 in. from each participant's mouth. The recordings were conducted in one session lasting approximately 45 min.

　　　Digital audio recordings were transferred to computer hard disk at a sampling rate of 22 kHz and 16-bit quantization. $F1$ and $F2$ of the vowel /ɛ/ were estimated for a total of 720 productions [10 talkers $\times$ 3 sentences $\times$ 12 repetitions $\times$ 2 conditions] at the midpoint of the steady-state portions of the vowel using Linear Predictive Coding (LPC) spectra combined with a peak-picking algorithm (Mertus, 2002). For each talker, there were a few cases where formant frequency estimation was difficult. For these cases, several steps were taken to ensure that formant frequency estimation was as accurate as possible including (a) centering a 20-ms hamming window at 20 and 40 ms to the left or right of the vowel midpoint and (b) identifying merged or missing formants through visual inspection of LPC spectra overlaid with discrete Fourier transform spectra. For each talker, the length of auditory deprivation was measured as the duration from the sentence onset to the midpoint of target vowel for all three sentence types produced in the processor-off condition (see bottom arrow in Fig. 1). This represents the average time for which auditory feedback was absent prior to vowel midpoint.

　　　Two graduate students carried out narrow phonetic transcriptions of the target word "head" produced in both processor-on and -off conditions. Transcribers heard the target word in the original sentence contexts via headphones. Transcribers were trained in phonetic transcription. They were monolingual, native speakers of American English and were blind to the results of the acoustic analyses. In cases of disagreement between the two transcribers, a third transcriber was enlisted to serve as a tie-breaker.

2.2 Results

In an effort to compensate for differences in vocal tract size, formant frequency data were normalized using Nearey's log mean normalization procedure (Nearey, 1989). A two-way analysis of variance was conducted (with processor conditions and sentence type as factors) using log-normalized formant frequency values. In addition, three planned comparisons (with *Bonferroni* corrections and family-wise *p* value set at $p < 0.0167$) were performed for each talker to determine whether there were any significant processor-on versus -off differences, across the three
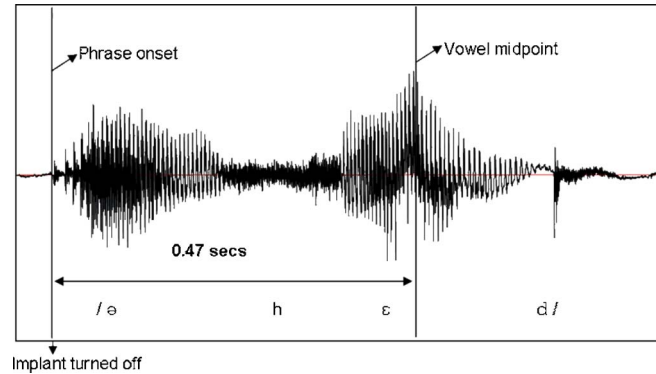
Fig. 1. (Color online) Waveform of the sentence "*a head*". The arrow at the bottom shows that auditory feedback is absent for approximately 0.47 s by the time the speaker has reached the midpoint of the target vowel /ɛ/.

sentence types. The F-ratios and Cohen's d values (Cohen, 1988) for each of the significant comparisons are reported in Table 2.

Figure 2 shows $F1$ and $F2$ values in kHz for the vowel /ɛ/ produced by ten talkers in three sentence contexts in both processor-on and -off conditions. The duration of auditory feedback deprivation associated with each sentence and each talker is listed above the subplots in Fig. 2. Figure 3 shows that there was no overlap in the length of auditory feedback deprivation across the three sentence types for any given talker. For sentence 1, with an average (across 12 repetitions) of $0.3-0.7$ s of auditory deprivation, one talker (T1) showed a statistically significant $F1$ decrease and two talkers (T6 and T8) showed significant $F2$ decrease. For sentence 2, with an average of $1.5-3.0$ s of auditory deprivation, three talkers (T6, T7, and T8) showed a statistically significant $F2$ decrease. None of the ten talkers showed a reliable $F1$ change. For sentence 3, with an average of $2.6-4.7$ s of auditory deprivation, one talker (T2) showed a statistically significant $F1$ decrease and two talkers (T6 and T8) showed a significant $F1$ increase. In addition, two talkers (T7 and T8) showed a significant $F2$ decrease. Talkers T3, T4, T5, T9, and T10 did not show any reliable changes in $F1$ or 2 for any sentence. The processor-on versus -off (absolute) differences that were significant ranged from 47.5 to 85.5 Hz for $F1$ and ranged from 94 to 278 Hz for $F2$.

Correlation analyses were performed to examine the relationship between the extent of absolute $F1$ and/or $F2$ changes in processor-off conditions and the length of auditory feedback deprivation across all three sentence types. For $F1$, a moderate, positive correlation was found for sentence 1 ($r=0.65$; $p<0.001$; $n=10$) and sentence 2 ($r=0.75$; $p<0.001$; $n=10$). For $F2$, a moderate, positive correlation was found for sentence 1 ($r=0.55$; $p<0.001$; $n=10$). None of the remaining correlations were significant. Correlation analyses were also conducted to examine the relationship between the extent of $F1$ and/or $F2$ changes in processor-off conditions and (a)

Table 2. F ratios and Cohen's d values for talkers who showed significant processor-on versus -off differences for $F1$ and $F2$ for the three sentence types, as confirmed by planned comparisons with Bonferroni corrections ($p<0.0167$).

S1	$F1$	T1 $[F(1,22)=12.1; d=1.5]$
	$F2$	T6 $[F(1,22)=10.9; d=1.0]$; T8 $[F(1,22)=8.82; d=1.3]$
S2	$F1$	*ns*
	$F2$	T6 $[F(1,22)=12.7; d=1.1]$; T7 $[F(1,22)=15.3; d=1.3]$; T8 $[F(1,22)=21.7; d=1.3]$
S3	$F1$	T2 $[F(1,22)=16.1; d=1.3]$; T6 $[F(1,22)=14.1; d=1.0]$; T8 $[F(1,22)=9.16; d=0.7]$
	$F2$	T7 $[F(1,22)=26.1; d=1.9]$; T8 $[F(1,22)=46.4; d=2.5]$
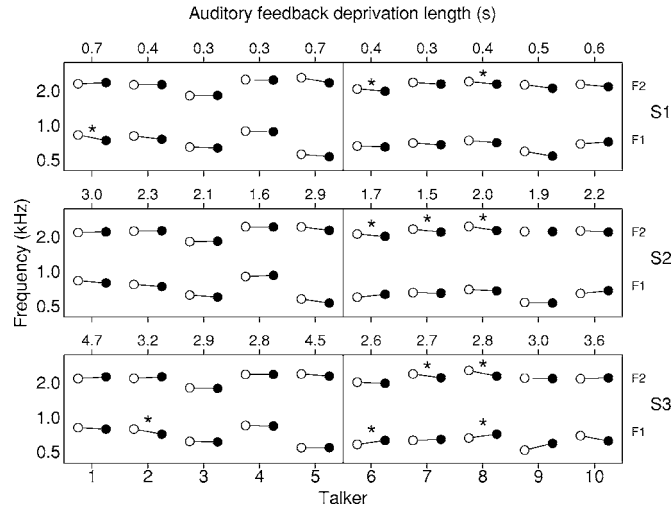
Fig. 2. Mean $F1$ and $F2$ values for vowel /ɛ/ produced in three sentences (S1, S2, and S3) by talkers 1–10 in processor-on and processor-off conditions. The average duration of auditory feedback deprivation in seconds is represented at the top of the dot plot for each talker. A star [*] indicates a significant processor-on versus -off difference as confirmed by planned comparisons with Bonferroni corrections ($p < 0.0167$).

the age of implantation and (b) the length of hearing experience for all three sentence types. A moderate, negative correlation was found between $F1$ and the age of implantation for sentence 2 ($r = -0.67$; $p < 0.001$; $n = 10$). None of the remaining correlations were significant.

A subset of the phonetic transcription data was examined for the five talkers (T1, T2, T6, T7, and T8) who showed significant processor-on versus -off difference in order to explore whether the measured acoustic effects were perceptible to trained listeners. Narrow transcriptions by the two transcribers were compared to evaluate the extent of agreement between them. Of the 216 comparisons, there were 21 instances of disagreement, which were resolved by a
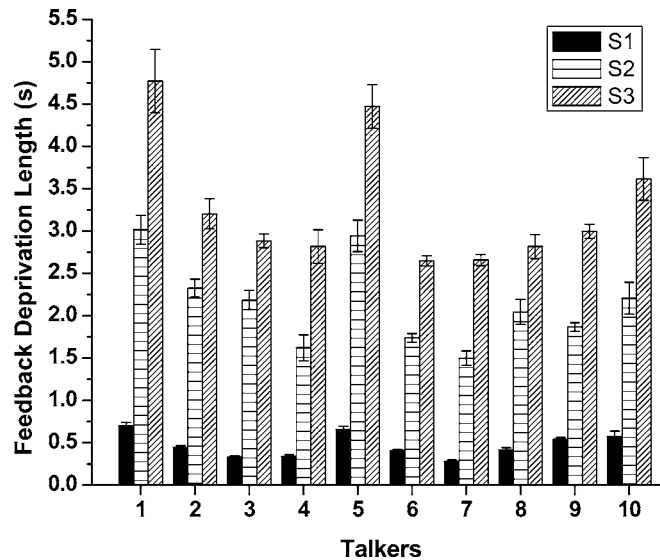
Fig. 3. Mean auditory feedback deprivation length (s) computed from sentence onset to vowel midpoint for sentences (S1, S2, and S3) produced by talkers 1–10 in processor-off conditions.

third transcriber. For sentence 1 produced by talker T1, phonetic transcription by both transcribers revealed more instances of substitution of /ɛ/ with /æ/ in processor-on (approximately 75% of the time) compared to the processor-off condition (approximately 50%). These substitutions are consistent with the acoustic findings of lowered $F1$ in processor-off compared to the processor-on condition. For talker T8, phonetic transcription by both transcribers showed several instances (average of 22%) of substitution of /ɛ/ with /æ/ in the processor-off condition only. This is also consistent with the acoustic findings of decrease in vowel $F2$ and an increase in vowel $F1$ in the processor-off condition. For talker T2, phonetic transcription by both transcribers revealed more instances of nasalized vowels in processor-off (approximately 91%) compared to the processor-on (approximately 72%) condition. The increase in nasality may be related to the decrease in vowel $F1$ in the processor-off condition. Lastly, for talkers T6 and T7, phonetic transcription by both transcribers did not reveal any noteworthy differences in vowels produced in the processor-on versus -off conditions.

3. General discussion

Auditory feedback deprivation resulted in $F1$ and $F2$ changes in the vowel /ɛ/ for a subset of talkers. While $F1$ shifts were characterized by both increases and decreases, significant $F2$ shifts were always lower in the absence of auditory feedback, suggesting vowel centralization. Centralization of vowels refers to the notion that vowels tend to drift toward the neutral vowel or middle of the vowel space due to impoverished feedback, leading to a constricted vowel space in some deaf individuals (Monsen, 1976). The vowel /ɛ/ is typically characterized by a higher $F2$ and a slightly lower $F1$ frequency when compared to a neutral vowel. Thus, a decrease in $F2$ and an increase in $F1$ for vowel /ɛ/ would be evidence of centralization. Five of the ten talkers showed significant $F1$ and/or $F2$ shifts when auditory feedback was removed. For those talkers who showed the largest acoustic effects for $F1$ and $F2$, phonetic transcription data revealed perceptible phonetic differences between the processor-on versus -off conditions.

Correlation analyses showed a positive relationship between the extent of formant frequency change and the duration of auditory deprivation, suggesting that larger changes in speech occurred for longer durations of auditory feedback deprivation. In addition, correlation analyses showed that the extent of $F1$ change was inversely related to the age of implantation, suggesting that children who were implanted at later ages showed smaller $F1$ changes in the absence of auditory feedback compared to children who received implants at younger ages.

Auditory deprivation experienced by children with hearing impairments could lead to limitations in terms of developing accurate representations between articulatory movements and their acoustic outcomes. Thus, responses to altered feedback may be different in children with hearing impairments compared to individuals with normal hearing. While it is not practical to have a control group for experiments using an on-off paradigm, the data from the present study showed that the responses to altered feedback by some children aided with cochlear implants were similar to those of normal-hearing adults. That is, consistent with the findings of Purcell and Munhall (2006a), three talkers in the present study showed significant shifts in $F1$ and/or $F2$ when auditory deprivation lasted approximately between $0.4-0.7$ s. One might hypothesize that the talkers for whom rehabilitation efforts have been successful are likely to demonstrate auditory-feedback-based corrective changes in their speech similar to those of normal-hearing individuals. In fact, all the children who showed significant acoustic effects demonstrated good speech perception scores, had implant experience of approximately 6–8 years, and demonstrated relatively high average speech intelligibility scores compared to the average speech intelligibility scores for the other five talkers who did not show any significant $F1$ or $F2$ shifts. It is possible that deaf children with good speech perception abilities and intelligible speech are able to use feedback and feedforward controllers efficiently and interactively (see Max *et al.*, 2004) to fine-tune their speech in a manner similar to normal-hearing adults.

In conclusion, vowel formant frequency shifts were noted in five talkers in the absence of auditory feedback. Three of the five talkers showed $F1$ and/or $F2$ shifts when auditory feedback was eliminated for as briefly as $400-700$ ms, suggesting that the time involved in implementing auditory feedback-based corrective changes in speech produced by some deaf children

is similar to that of normal-hearing adults. Data from the present study provides evidence that auditory deprivation can produce small but perceptible acoustic changes in vowels produced by some pediatric cochlear-implant users. Future investigations are clearly necessary to extend the present study to a larger set of acoustic measures involving consonants as well as vowels.

Acknowledgments

References and links

Bharadwaj, S. V., Tobey, E. A., Assmann, P. F., and Katz, W. F. (**2006**). "Effects of auditory feedback on fricatives produced by cochlear-implanted adults and children: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **119**, 1626–1635.

Borden, G. (**1979**). "An interpretation of research on feedback interruption during speech," Brain Lang **7**, 302–319.

Cohen, J. (**1988**). *Statistical power analysis for the behavioral sciences* (Lawrence Erlbaum, Hillsdale, NJ).

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of frequency-shifted auditory feedback on voice F0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

Guenther, F. H. (**2006**). "Cortical interactions underlying the production of speech sounds," J. Commun. Dis. **39**, 350–365.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Kawahara, H., and Williams, J. C. (**1996**). "Effects of auditory feedback on voice pitch trajectories: Characteristic responses to pitch perturbations," in *Vocal Fold Physiology: Controlling complexity and chaos*, edited by P. J. Davis and N. H. Fletcher (Singular Publishing Group, San Diego, CA), pp. 263–278.

Matthies, M. L., Svirsky, M., Perkell, J., and Lane, H. (**1996**). "Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant," J. Speech Hear. Res. **39**, 936–946.

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., and Wallace, M. E. (**2004**). "Unstable or insufficiently activated internal models and Feedback-Biased motor control as sources of dysfluency: A theoretical model of stuttering," Contemporary Issues in Comm. Sci. Dis. **31**, 105–122.

Mertus, J. (**2002**). Brown Lab Interactive Speech System (BLISS) Software.

Monsen, R. B. (**1976**). "The production of English stop consonants in the speech of deaf children," J. Phonetics **4**, 29–41.

Nearey, T. M. (**1989**). "Static, dynamic and relational properties in speech perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Perkell, J. S., Lane, H., Svirsky, M. A., and Webster, J. (**1992**). "Speech of cochlear implant patients: A longitudinal study of vowel production," J. Acoust. Soc. Am. **91**, 2961–2979.

Perkell, J., Guenther, F. H., Lane, H., Matthies, M., Perrier, P., Vick, J., Wilhems-Tricario, R., and Zandipour, M. (**2000**). "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," J. Phonetics **28**, 233–272.

Purcell, D. W., and Munhall, K. G. (**2006a**). "Compensation following real-time manipulation of formants in isolated vowels," J. Acoust. Soc. Am. **119**, 2288–2297.

Purcell, D. W., and Munhall, K. G. (**2006b**). "Adaptive control of vowel formant frequency: evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.

Svirsky, M. A., and Tobey, E. A. (**1991**). "Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users," J. Acoust. Soc. Am. **89**, 2895–2904.

Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (**1992**). "Effects of short-term deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–1300.

Tartter, V. C., Chute, P. M., and Hellman, S. A. (**1989**). "The speech of a postlingually deafened teenager during the first year of use of a multichannel cochlear implant," J. Acoust. Soc. Am. **86**, 2113–2121.

Tobey, E. A., Geers, A. E., Brenner, C., Altuna, D., and Gabbert, G. (**2003**). "Factors associated with development of speech production skills in children implanted before age five," Ear Hear. **24**, 36S–45S.

Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Zimmermann, G., and Rettaliata, P. (**1981**). "Articulatory patterns of an adventitiously deaf speaker: Implications for the role of auditory information in speech production," J. Speech Lang. Hear. Res. **24**, 169–178.

An exact point source starting field for the Fourier parabolic equation in outdoor sound propagation

Kenneth E. Gilbert and Xiao Di
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677
kgilbert@olemiss.edu, xiaodi@olemiss.edu

Abstract: A method for exactly representing a point source starting field in a Fourier parabolic equation calculation is presented. The formulation is based on an exact, analytic expression for the field in vertical wave number space (k space). The field in vertical coordinate space (z space) is obtained via a Fourier transform of the k-space field. Thus, one can directly control the Fourier components of the starting field, so that nonpropagating components are excluded. The relation of the exact starting field to the standard Gaussian starting field is demonstrated analytically. Examples of the numerical implementation of the exact starting field are given.

1. Introduction

A parabolic equation calculation typically begins with a numerical statement of the "starting field," which is the complex pressure as a function of height at the initial range, generally taken to be zero. The starting field originally proposed by Tappert was a Gaussian function of height that evolved into an approximation for the field of a point source.[1] Numerous improvements to the Gaussian starting field have been proposed, both in ocean acoustics and atmospheric acoustics.[2,3] Although various forms of the Gaussian starting field have been successfully used for many years, when the source is on the ground surface, problems arise due to the finite vertical extent of the Gaussian.[3,4] We present here a method for representing a true point source starting field in a Fourier parabolic equation calculation. The method is essentially exact and is valid even when the source height is zero.

In attempting to specify an exact starting field for a point source, one obvious difficulty is that the field diverges at the location of the source. Consequently, no exact finite starting field at zero range ($r=0$) can be defined in vertical coordinate space (z space). The essential idea here is that we do not have to specify the starting field in z space at $r=0$, because we can specify it one range step away ($r=\Delta r$). This approach can be applied directly in z space, as in Ref. 4, assuming one has an exact solution for a point source. For a finite impedance surface, however, no analytic z-space solution exists, so that one must compute the exact z-space field numerically. Then the numerical solution in z space must be numerically Fourier transformed to obtain a "filtered" numerical solution in vertical wave number space (k space). With care, such an approach can give accurate results.[4] In contrast to the direct z-space approach, the approach presented here takes advantage of the fact that, in k space, one has an exact, analytic result, even for a finite impedance surface. Thus, we can simply begin with the k-space solution at $r=\Delta r$, thereby eliminating the need to first compute an exact numerical point source solution in z space and then Fourier transform it to obtain a filtered k-space solution.

Since the point source field in k space is exact and analytic at all ranges, a filtered point source field in z space at any nonzero range can be easily and efficiently computed from the k-space field by means of a fast Fourier transform (FFT). The k-space method thus uses the same computational procedure as does the Fourier parabolic equation itself. The advantages of the approach, which is similar to that proposed in Refs. 5 and 6, are threefold: first, the point

source solution is essentially exact; second, the *k*-space components can be controlled directly in the *k*-space FFT; and, third, there is no limitation on the source height, which can be taken to be at $z_0=0$, if needed. Although the formulation presented in this article can, in principle, be used for any parabolic equation calculation, we consider a specific Fourier parabolic equation method known as the "Green's function parabolic equation" or "GFPE".[7]

2. GFPE starting field

To understand what is involved in propagating the starting field from the origin to the end of the first range step in a GFPE calculation, it is helpful to consider how the starting field enters into the calculation.

The GFPE solves for $\Psi(r,z)=\sqrt{r}P(r,z)$, where P is the complex pressure in the frequency domain, and r and z are, respectively, horizontal and vertical distance. The calculation assumes a far-field wave equation and one-way (no backscatter) propagation. For the sake of simplicity, in everything presented here, the so-called "carrier wave" has not been removed.

For a constant sound speed, c_0, to obtain the field in z space on the first range step, i.e., $\Psi(\Delta r,z)$, one evaluates the following integral, which is a Fourier transform with respect to the vertical wave number:[7]

$$\Psi(\Delta r,z)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}[\tilde{\Psi}(0,k')+\mathcal{R}(k')\tilde{\Psi}(0,-k')]\times e^{i\sqrt{k_0^2-k'^2}\Delta r}e^{ik'z}dk'$$
$$+2ik_s\tilde{\Psi}(0,k_s)e^{i\sqrt{k_0^2-k_s^2}\Delta r}e^{-ik_sz}, \tag{1}$$

where $\mathcal{R}(k')$ is the plane-wave reflection coefficient, which can be written

$$\mathcal{R}(k')=\frac{k'-k_s}{k'+k_s}. \tag{2}$$

In Eqs. (1) and (2), the quantity k_s is defined as $k_s=k_0/\hat{Z}$, where $k_0=\omega/c_0$, and $\hat{Z}$ is the normalized ground impedance. The quantity $\tilde{\Psi}(0,k')$ in Eq. (1) is the Fourier transform of $\Psi(0,z)$, and is defined as

$$\tilde{\Psi}(0,k')=\int_{-\infty}^{+\infty}\Psi(0,z')e^{-ik'z'}dz'. \tag{3}$$

Note that to apply Eq. (1) and Eq. (3), one must specify a starting field $\Psi(0,z')$. A main point of the present paper is that even though $\Psi(0,z')$ does not exist as a finite mathematical entity for a point source, one can nevertheless easily incorporate an exact point source starting field into the standard GFPE formulation. The approach taken begins with the exact integral expression for the acoustic field of a point source.

3. Integral expression for a point source

We begin by outlining the solution for a point source in a constant sound-speed atmosphere over a surface that is represented by an angle-independent, frequency-dependent complex ground impedance.

For an atmosphere with constant sound speed, c_0, the frequency-domain solution for a point source is the Green's function for the Helmholtz equation, which satisfies

$$(\nabla^2+k_0^2)G(\boldsymbol{R},\boldsymbol{R}_0)=-4\pi\delta(\boldsymbol{R}-\boldsymbol{R}_0), \tag{4}$$

where $\boldsymbol{R}=(x,y,z)=(r,z)$, $\boldsymbol{R}_0$ is the source position, and $\delta(\boldsymbol{R}-\boldsymbol{R}_0)$ is a three-dimensional Dirac delta function. We write the Green's function in a well-known integral form[2]

$$G(\boldsymbol{R},\boldsymbol{R}_0) = \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\boldsymbol{\kappa}\cdot\boldsymbol{r}}g(\kappa,z,z_0)d\kappa_x d\kappa_y = \int_{-\infty}^{\infty} H_0^{(1)}(\kappa r)g(\kappa,z,z_0)\kappa d\kappa, \qquad (5)$$

where $\boldsymbol{\kappa}=(\kappa_x,\kappa_y)$ is the vector horizontal wave number, and $H_0^{(1)}$ is a Hankel function of the first kind. The one-dimensional Green's function $g(\kappa,z,z_0)$ in Eq. (5) satisfies

$$\left[\frac{\partial^2}{\partial z^2} + k_v^2\right]g(\kappa,z,z_0) = -\delta(z-z_0), \qquad (6)$$

where $k_v=\sqrt{k_0^2-\kappa^2}$ is the vertical wave number, and $\delta(z-z_0)$ is a one-dimensional Dirac delta function.

From this point on, the approach is the same as for the GFPE development in Ref. 7, except that we are solving for the field of a point source rather than advancing a starting field. Thus, the horizontal dependence is given by $H_0^{(1)}(\kappa r)$, as in Eq. (5), instead of $\exp(i\kappa r)$ as in Ref. 7.

The solution to Eq. (6) for $g(\kappa,z,z_0)$ is the same as with the GFPE, and is given by

$$g(\kappa,z,z_0) = \frac{i}{2k_v}[e^{ik_v|z-z_0|} + \mathcal{R}(k_v)e^{ik_v(z+z_0)}]. \qquad (7)$$

To proceed, we need to write Eq. (7) in terms of an integral over vertical wave number. By doing so, we can replace the horizontal wave number integral in Eq. (5) with a vertical wave number integral.

Cauchy's integral formula[8] can be used to arrive at two identities needed to express Eq. (7) as an integral over vertical wave number

$$\frac{i}{2k_v}e^{ik_v|z-z_0|} \equiv \int_{-\infty}^{\infty} \frac{e^{ik'(z-z_0)}}{k'^2 - k_v^2}\frac{dk'}{2\pi} \qquad (8)$$

and

$$\frac{i}{2k_v}\mathcal{R}(k_v)e^{ik_v(z+z_0)} \equiv \int_{-\infty}^{\infty} \mathcal{R}(k')\frac{e^{ik'(z+z_0)}}{k'^2 - k_v^2}\frac{dk'}{2\pi} + 2ik_s\frac{e^{-ik_s(z+z_0)}}{k_s^2 - k_v^2}. \qquad (9)$$

Equations (8) and (9), together with Eq. (7), yield the needed one-dimensional Green's function, which is then substituted into Eq. (5). In Eq. (5), the integration order is rearranged, so that the integration with respect to the horizontal wave number κ is done first

$$G(\boldsymbol{R},\boldsymbol{R}_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik'(z-z_0)}dk' \int_{-\infty}^{\infty} \frac{H_0^{(1)}(\kappa r)}{\kappa^2 - (k_0^2 - k'^2)}\kappa d\kappa$$

$$+ \frac{1}{2\pi}\int_{-\infty}^{\infty} R(k')e^{ik'(z+z_0)}dk' \int_{-\infty}^{\infty} \frac{H_0^{(1)}(\kappa r)}{\kappa^2 - (k_0^2 - k'^2)}\kappa d\kappa$$

$$+ 2ik_s e^{-ik_s(z+z_0)}\int_{-\infty}^{\infty} \frac{H_0^{(1)}(\kappa r)}{\kappa^2 - (k_0^2 - k_s^2)}\kappa d\kappa. \qquad (10)$$

The Hankel function $H_0^{(1)}(\kappa r)$ goes exponentially to zero in the upper half plane for $|\kappa r|\to\infty$, and is analytic everywhere except for a branch cut on the negative κ axis. Hence, contour integration can be used for the κ integral. The integration contour is the same as shown in Fig. 1 of Ref. 7, where the poles κ_+ and κ_- represent, respectively, the positive and negative values of $\sqrt{k_0^2-k'^2}$ or $\sqrt{k_0^2-k_s^2}$. We assume that k_0 has a small imaginary part so that κ_+ is always in the upper half plane and κ_- is always in the lower half plane. Using contour integration, we have

$$G(\boldsymbol{R},\boldsymbol{R}_0) = \frac{i}{2}\int_{-\infty}^{\infty} H_0^{(1)}(\sqrt{k_0^2-k'^2}r)e^{ik'(z-z_0)}dk' + \frac{i}{2}\int_{-\infty}^{\infty} H_0^{(1)}(\sqrt{k_0^2-k'^2}r)\mathcal{R}(k')e^{ik'(z+z_0)}dk'$$

$$- 2\pi k_s e^{-ik_s(z+z_0)}H_0^{(1)}(\sqrt{k_0^2-k_s^2}r). \tag{11}$$

The right-hand side of Eq. (11) is an exact point-source solution and can be interpreted as three distinct physical contributions: the first term is the direct wave; the second term is the reflected wave; and the third term is the surface wave. Note that the complex pole in the reflection coefficient in Eq. (2), $k'=-k_s$, is also the complex vertical wave number of the surface wave in Eq. (11).

   The integral in the first term of Eq. (11) is just the free-field point-source solution, $\exp(ik_0|\boldsymbol{R}_1|)/|\boldsymbol{R}_1|$, where $|\boldsymbol{R}_1| = \sqrt{r^2+(z-z_0)^2}$. However, we leave the first term in integral form, since it will be convenient later to numerically evaluate both the first and second terms in Eq. (11) as a single integral.

4. Incorporating the point source solution into the GFPE

Even though, in z space, the field for a point source is singular for $r=0$, it is possible, nevertheless, to use the point source solution in Eq. (11) to deduce an exact, analytic k-space starting field at any range. When Fourier transformed at $r=\Delta r$, the k-space field yields an exact point source field in z space at the end of the first range step.

   On the first range step in the GFPE, the Hankel functions in Eq. (11) are generally evaluated at a horizontal distance of greater than a wavelength. Thus, the large-argument asymptotic form can be used:[8] $H_0^{(1)}(\sqrt{k_0^2-k'^2}r) \approx \sqrt{2/i\pi r}\exp(i\sqrt{k_0^2-k'^2}r)/(k_0^2-k'^2)^{1/4}$. Using the asymptotic form and writing the first two terms as a single Fourier transform with respect to vertical wave number, we can put Eq. (11) into a form easily used in the GFPE. Since the GFPE computes a complex amplitude Ψ that is $\sqrt{r}$ times the complex pressure, Eq. (11) becomes for the first range step (assumed to be greater than a wavelength)

$$\Psi(\Delta r, z) = \sqrt{\frac{i}{2\pi k_0}}\int_{-\infty}^{\infty} [e^{-ik'z_0} + \mathcal{R}(k')e^{ik'z_0}]\frac{e^{i\sqrt{k_0^2-k'^2}\Delta r}}{(1-k'^2/k_0^2)^{1/4}}e^{ik'z}dk'$$

$$- \sqrt{\frac{2\pi}{ik_0}}2k_s e^{-ik_s(z+z_0)}\frac{e^{i\sqrt{k_0^2-k_s^2}\Delta r}}{(1-k_s^2/k_0^2)^{1/4}}. \tag{12}$$

Within the accuracy of the asymptotic form for the Hankel function, the expression in Eq. (12) gives an exact result for the acoustic field at a horizontal distance of Δr.

   One advantage of the point-source solution in Eq. (12) is that the vertical wave number components are controlled by the limits in the k' integration. To obtain a numerically accurate result for a short initial range (e.g., a fraction of a wavelength), one must integrate over a range exceeding the interval $(-k_0, k_0)$. For example, we could integrate over the interval $(-2k_0, 2k_0)$. For practical GFPE calculations, however, where the first range step is typically greater than a wavelength, the Fourier components with $|k'| > k_0$ decay rapidly horizontally. Hence, for these calculations, the acoustic field contains negligible vertical wave number components greater than $\pm k_0$. Since a plane wave with vertical wave number $\pm k_0$ propagates directly upward/downward, it is physically clear why the wave number components $|k'| \geq k_0$ are not important in practical calculations. In fact, such calculations typically use a k-space filter that rolls off to zero before reaching $\pm k_0$.[9] In Sec. 6, numerical examples will be given demonstrating how the k-space integration affects the starting field.

   Finally, if the sound speed is not a constant, but rather $c(r,z)$, one would multiply the result in Eq. (12), by the "phase correction" $e^{i[k(\Delta r/2,z)-k_0]}$, where $k(\Delta r/2,z) = \omega/c(\Delta r/2,z)$.

5. Relation of the exact starting field to the standard Gaussian starting field

In this section, we show the connection between the exact starting field discussed above and the standard Gaussian starting field. The purpose is not to obtain an improved Gaussian starting field, but rather, simply to show that a connection exists.

First, the GFPE requires that we specify $\Psi(0,z)$, but if we try to obtain a result for $\Psi(0,z)$ by setting $\Delta r = 0$ in Eq. (12), we find that the integral does not converge. Similarly, the integral does not exist with the exact Hankel function in Eq. (11). Hence, as discussed earlier, no exact, finite result for Ψ exists at $r=0$. With some approximations, however, a well defined, but approximate, result for $\Psi(0,z)$ can be derived and put into the same form as the standard Gaussian starting field.

First, in Eq. (12), we approximate the fourth roots in the denominators as unity. This approximation corresponds to a "narrow-angle" approximation, since it is valid for the part of the acoustic field with vertical wave number components that are much less that k_0. With $\Delta r = 0$, the integral involving $\exp[ik'(z-z_0)]$ yields $2\pi\delta(z-z_0)$. The integral with $\mathcal{R}(k')\exp[ik'(z,z_0)]$ cancels the surface wave term. Hence we have, $\Psi(0,z)=\sqrt{2\pi i/k_0}\,\delta(z-z_0)$. The Dirac delta function $\delta(z-z_0)$ can be approximated with a Gaussian of the form $\Gamma(z,z_0)=\exp[-(z-z_0)^2/L^2]/L\sqrt{\pi}$, where L is small relative to a wavelength. With these approximations, the starting field becomes the standard Gaussian starting field, $\Psi(0,z)=\sqrt{2i\pi/k_0}\,\Gamma(z,z_0)$. It is worth noting that a Gaussian approximation to $\delta(z-z_0)$ is not unique. A Dirac delta function can be approximated with any peaked function that is narrow compared to a wavelength and that integrates to unity. Thus, even with the above approximations, $\Psi(0,z)$ is not a unique function.

Although Gaussian starting functions have been used successfully for many years in both ocean and atmospheric acoustics, a practical problem arises when z_0 is small relative to L. For small values of z_0, $\Gamma(z,z_0)$ "overlaps" the surface at $z=0$, and part of the starting field is lost. An *ad hoc* solution has been to add an "image" source $\mathcal{R}(k_0)\sqrt{2i\pi/k_0}\,\Gamma(z,-z_0)$, where $\mathcal{R}(k_0)$ is the normal incidence reflection coefficient.[3,7] The purpose of the image source is to approximately fit the surface boundary conditions and thus to approximately recover the source strength lost by the overlap of the Gaussian source with the surface at $z=0$. The overlap problem does not occur with Eq. (12), because we have the field for a point source for any height $z \geq 0$. In fact, with a Gaussian starter, it appears that using a delta function (i.e., letting $L \to 0$) instead of the standard Gaussian, would be a better approximation to the exact starting field, since there would be no problem with overlap.

6. Numerical implementation of the exact starting field

The numerical implementation is no more involved than for the standard Gaussian starting field. However, because of the k-space integration, which is done numerically with a FFT, some computational issues need to be addressed.

We first show that, when evaluated numerically over a sufficiently large k-space interval and with a sufficiently fine numerical mesh, Eq. (12) yields an essentially exact result. Figures 1(a) and 1(b) show a comparison of the relative sound pressure level versus range and height (i.e., normalized to spherical spreading), for an exact Laplace transform method[10] and for Eq. (12). The source height is 1.8 m, the frequency is 500 Hz, and the normalized ground impedance is $7.19+8.2i$.

To obtain an accurate result at very short ranges (e.g., less than a wavelength) using Eq. (12), it is necessary to include nonpropagating vertical wave number components. Hence the integral must go beyond $\pm k_0$, (e.g., $-2k_0$ to $2k_0$). The fourth root factor is singular at $|k'|=k_0$, but is nevertheless integrable in terms of a hypergeometric function.[11] In principle, such integrable singularities should be treated by "subtracting" the singularity.[12] In practice, it often suffices to use a very fine numerical mesh and avoid placing a point exactly on the singularity. The second approach was used here to produce Fig. 1(b). Although the method used is not efficient, it is sufficiently accurate to demonstrate that Eq. (12) produces an exact result if evalu-
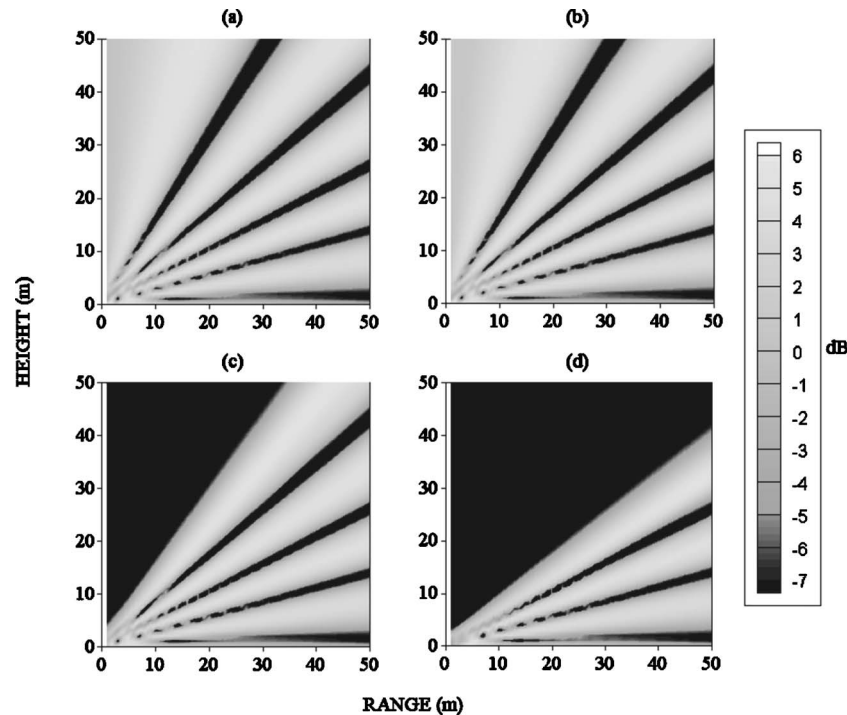
Fig. 1. Relative sound pressure level versus range and height for a point source at a height of 1.8 m. The frequency is 500 Hz, and the normalized ground impedance is $7.19+8.2i$. Figure 1(a) is an exact calculation using a Laplace transform method, and Fig. 1(b) is an evaluation of Eq. (12) with the integral going from $-2k_0$ to $+2k_0$. The integrals in 1(c) and 1(d) go from $-k_0$ to $+k_0$ and $-0.8k_0$ to $+0.8k_0$, respectively, with a smooth roll-off over a $0.2k_0$ interval at the ends of the integrals. In Figs. 1(c) and 1(d), because of the roll-off filter in the integrals, the acoustic field has an aperture of less than 90°.

ated accurately. The results shown in Figs. 1(a) and 1(b) start at a range of 1 m, so that the highest angle shown is 88.9°. At this short range and extreme angle, the error is 1.5 dB, but decreases rapidly with increasing range. At a more typical initial range, the error is considerably less, even at very high propagation angles. For example, at ranges of 3 and 5 m, respectively, the maximum angles are 86.6° and 84.3°, and the corresponding errors are 0.17 and 0.034 dB. Thus, with a sufficiently large integration interval and a sufficiently fine numerical mesh, Eq. (12) yields an essentially exact starting field.

For an intial range step of a wavelength or more, the factor $\exp(i\sqrt{k_0^2-k'^2}\Delta r)$ decays rapidly for k'^2 values exceeding k_0^2, thereby eliminating essentially all the vertical wave number components greater than k_0. Further, for the same reason, practical GFPE calculations typically integrate over a k-space interval less than $(-k_0, k_0)$.[9] Consequently, in practical calculations, where the first and succeeding range steps are generally at least several wavelengths, only propagating components ($|k'| \le k_0$) must be included. Physically, having $|k'| < k_0$ means that the maximum propagation angle is less than 90° with respect to horizontal. Examples of such calculations are shown in Figs. 1(c) and 1(d), where the integral is over the interval $\pm k_0$ and $\pm 0.8k_0$, respectively. A k-space filter was used at the ends of the integral to make the integrand go smoothly to zero before reaching the singularities at $\pm k_0$. In both cases, near the ends of the integration interval, the filter rolls off from unity to zero over a distance of $0.2k_0$. Because of the filter, the integrals in Figs. 1(c) and 1(d) are nonsingular, but do not include the very high-angle wave number components. As a result, the acoustic starting fields have large but finite apertures. The apertures in Figs. 1(c) and 1(d) are approximately 57° and 39°, respectively, and thus are applicable to typical GFPE calculations where propagation angles rarely exceed about 20°.
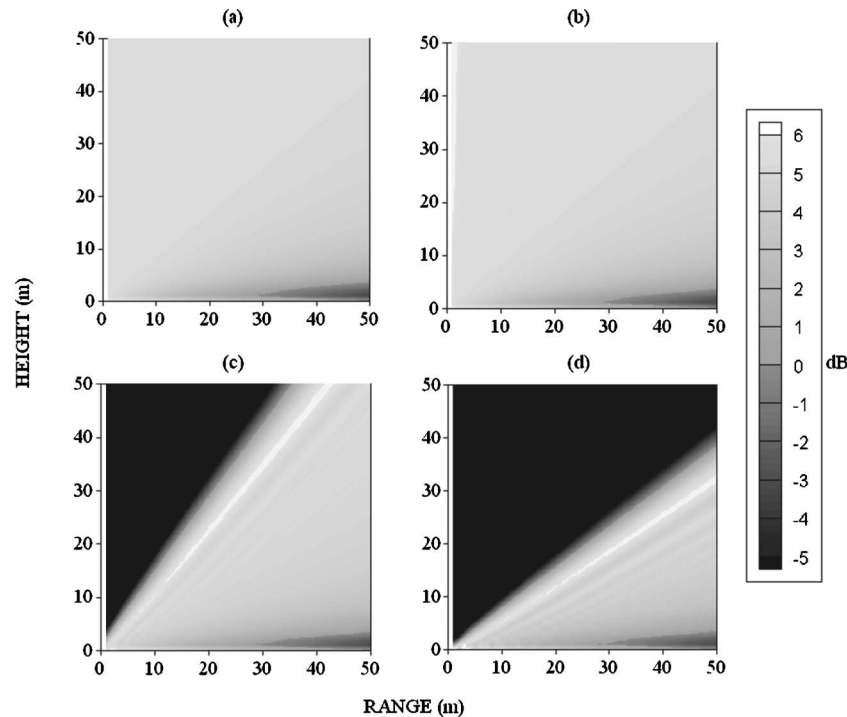
Fig. 2. As in Fig. 1, except the source is on the surface ($z_0=0$). Because the source and image source have zero separation, the dipole pattern seen in Fig. 1, is absent in Fig. 2. In addition to producing a reduced aperture, the roll-off filter in Figs. 1(c) and 1(d) causes some ripples at higher propagation angles.

Figure 2 shows the same calculation as in Fig. 1, except the source is on the surface ($z_0=0$). Note that the dipole pattern seen in Fig. 1 is absent in Fig. 2. The dipole pattern results from constructive and destructive interference between the source and its image. The smaller the source height, the larger the interference lobes. At zero height, the lobes are infinitely wide, so that a uniform relative sound pressure level results. In addition to the finite apertures as seen in Figs. 1(c) and 1(d), some edge effects (ripples) are evident in Figs. 2(c) and 2(d). Even with the edge effects, the starting field in Fig. 1(c) provides a useful aperture exceeding 45°.

For extreme high-angle propagation (e.g., for nearly overhead airborne receivers), the k-space integration, including the roll-off filters, should extend beyond the interval $(-k_0, k_0)$, so that a full 90° aperture is obtained. Such applications are not common, but nevertheless justify a more efficient treatment of the integrable singularities at $|k'|=k_0$ than was done here. Methods for efficiently subtracting the integrable singularity are presently being investigated.

7. Summary and conclusions

A method has been derived for computing an exact point source starting field for parabolic equation calculations in outdoor sound propagation. An exact analytic expression is given for the acoustic field in k space at any range. A Fourier transform (with a FFT) of the k-space field yields the z-space field at $r=\Delta r$. Thus, the Fourier components of the starting field can be directly controlled through the k-space Fourier transform. The accuracy of the formulation does not depend on the source height, which can be at $z_0=0$, if necessary. The method should be useful in applications, such as surface explosions and tire noise, where the source is on or very near a finite impedance ground surface.

Acknowledgment

References and links

[1] F. D. Tappert, "The parabolic approximation method," in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, New York, 1977), Chap. V, pp. 224–287.

[2] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).

[3] E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer, Boston, 2001).

[4] J. L. Cooper and D. C. Swanson, "Parameter selection in the Green's function parabolic equation," Appl. Acoust. **68**, 390–402 (2007).

[5] D. J. Thomson and C. S. Bohun, "A wide-angle initial field for parabolic equation models," J. Acoust. Soc. Am. **183**, S118 (1988).

[6] D. J. Thomson, "Wide-angle parabolic equation solutions to two range-dependent benchmark problems," J. Acoust. Soc. Am. **87**, 1514–1520 (1990).

[7] K. E. Gilbert and X. Di, "A fast Green's function method for one-way sound propagation in the atmosphere," J. Acoust. Soc. Am. **94**, 2343–2352 (1993).

[8] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953).

[9] E. M. Salomons, "Improved Green's function parabolic equation method for atmospheric sound propagation," J. Acoust. Soc. Am. **104**, 100–111 (1998).

[10] X. Di and K. E. Gilbert, "An exact Laplace transform formulation for a point source above a ground surface," J. Acoust. Soc. Am. **93**, 714–720 (1993).

[11] M. Abramowitz and I. A. Stegun, eds. *Handbook of Mathematical Functions* (U.S. Government Printing Office, Washington, D.C., 1968), p. 558.

[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992), pp. 788–789.

Destabilization of velocity feedback controllers with stroke limited inertial actuators

Oliver Nicholas Baumann and Stephen John Elliott

Institute of Sound and Vibration Research, University of Southampton, University Road, Southampton, Hampshire SO17 1BJ United Kingdom

onb@isvr.soton.ac.uk, sje@isvr.soton.ac.uk

Abstract: It has been observed when using inertial actuators for the active reduction of structural vibration, that velocity feedback controllers are liable to become unstable if the actuator is subject to stroke saturation. This article presents a simple nonlinear, time domain model of an inertial actuator mounted on a single degree of freedom system. At low amplitudes the actuator, when used in a velocity feedback control loop, increases the effective damping of the structure. At higher amplitudes the system is shown to become unstable, however, and generates limit cycle oscillations having a predictable frequency.

1. Introduction

Velocity feedback control, in which the velocity of a structure is fed back to a collocated force actuator, is a well-known method of increasing damping and, hence, reducing the vibration of the structure on which it acts.[1] The use of inertial, or proof mass, actuators as the forcing device in such controllers has also been well documented.[2–5] Figure 1 shows the cross section of such a device comprising a proof mass, which in this case also acts as the magnet, a conducting coil, and a casing. The device is attached to the structure to which one wishes to apply a force. The current in the coil serves to produce a force which acts between the actuator proof mass and the casing and, thus, the structure.

In any practical actuator design, the stroke of the proof mass is limited by the dimensions of the casing or end-stops which are incorporated in the design. When the actuator is subject to large or prolonged inputs, the proof mass will hit the end-stops and, hence, the displacement will saturate, resulting in an impulse which is transmitted to the structure.

The issue of stroke saturation has been addressed by Chase *et al.*[6] in which the authors observe that stroke saturation can lead to destabilizing forces in the control of a buckling beam and performance degradation in vibration control. The authors propose a solution involving the implementation of a secondary controller which feeds back the relative displacement and velocity of the proof mass, a concept which has seen more attention recently.[5] It should be noted, however, that the system under control in this case, namely, the buckling beam, was inherently unstable and the purpose of control was to stabilize the system. Earlier work[7,8] investigated both linear and nonlinear controllers, which are reported to reduce the tendency of the actuator proof mass to collide with its end-stops, although the cited motivation for this work was to prevent damage to the system and no mention was made of the system's stability. It has been observed by the authors, however, that, when velocity feedback loops are used in practice to control structural vibration, the forces due to the displacement saturation of the proof mass can destabilize the system. This problem is particularly severe in a system with multiple actuators and local velocity controllers, since instability in one loop has been observed to induce instability in all other loops.[9] This work concentrates on the modelling of a stroke limited inertial actuator used in a velocity feedback control regime on a simple, but inherently stable, dynamic system. It is observed that for large disturbances that result in control inputs which cause the actuator to saturate, the modelled system becomes unstable, as has been observed experimentally.
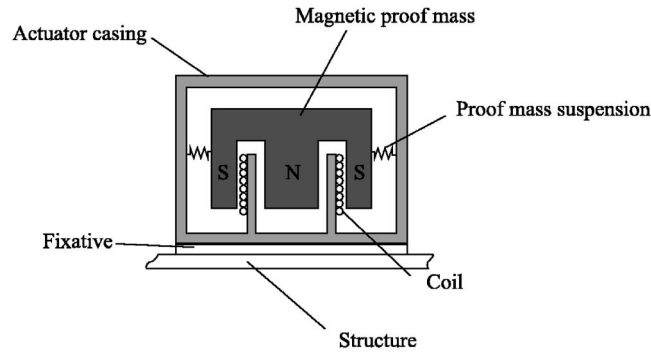
Fig. 1. A schematic of an inertial actuator in cross section.

In this paper we present a lumped parameter model of an inertial actuator on a structure modelled as a single degree of freedom. The finite stroke of the actuator proof mass is modelled as a nonlinear stiffness. The actuator is then driven using a current proportional to the velocity of the structure. The response of the system to an impulsive force is presented in the form of time series data and shown to become unstable for large impulse magnitudes.

2. The lumped parameter model

The actuator, attached to a single degree of freedom structure is modelled using a lumped parameter system as in Fig. 2. In practice the single degree of freedom structure would typically model the lowest structural resonance. The actuator proof mass m_p is, therefore, coupled to the structural mass m_s via the stiffness and damping parameters associated with its suspension, $k_p(w)$ and c_p, respectively, where $k_p(w)$ is dependant on the relative displacement of the inertial mass and the structure $w(t)$. This allows the displacement saturation of the proof mass to be modelled as the suspension system taking a very large stiffness. The mass associated with the structure is connected to an inertial reference via the stiffness and damping parameters k_s and c_s, respectively. The displacements of the proof mass and structure are defined as $w_p(t)$ and $w_s(t)$, respectively, so that the relative displacement $w(t)=w_p(t)-w_s(t)$. The primary force $f_p(t)$
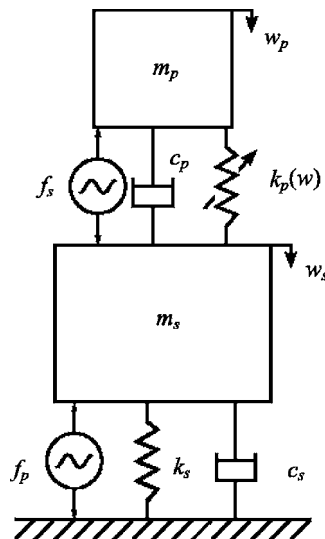
Fig. 2. A lumped parameter model of the actuator and structure.

acts between the structure and the inertial reference and the secondary force due to the current in the actuator coil $f_s(t)$ acts between the proof mass and the structure. To avoid numerical problems associated with parameter discontinuities, we define the stiffness to be a continuous function of the form

$$k_p(w) = k_{\text{lin}}\left(1 + \frac{(w(t))^{N-1}}{w_0^N}\right),$$ (1)

where k_{lin} is the stiffness of the proof mass suspension when the proof mass is in its linear range, being between $-w_0$ and w_0, and N is an odd integer related to the rate at which the stiffness of the suspension increases as the proof mass approaches saturation. In this way, as the actuator proof mass approaches the limits of its linear range, it is prevented from moving any further, relative to the structure, by the increase in stiffness. The secondary force, acting between the structure and the proof mass, then becomes progressively less effective. This relatively simple model of the proof mass saturation has been adopted to ensure that both momentum and energy are conserved in the collision of the proof mass and its end stops. The variable N may be altered to approximately model different types of end-stop, a hard end-stop corresponding to a high value, $N=49$ is used in the simulations below, and a soft end-stop to a lower value. When the actuator proof mass is small compared with w_0, the stiffness is constant and the relative displacement varies linearly with the applied force. As the relative displacement approaches $\pm w_0$, the proof mass makes contact with the end-stops and the force required to appropriate further extension increases significantly. The dynamics of system described above may be expressed in state space form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_p f_p(t) + \mathbf{B}_s f_s(t),$$ (2)

where $\mathbf{x}$ is the state vector comprising the displacements and velocities of the proof and structural masses, $\mathbf{A}$ is the system matrix describing the coupled dynamics of the masses, $\mathbf{B}_p$ and $\mathbf{B}_s$ describe the relationship between the system states and the primary and secondary forces, respectively.

We define the output of the system y to be the velocity of the structure mass which is related to the state vector via the output equation

$$y(t) = \mathbf{C}\mathbf{x}(t).$$ (3)

With fixed gain, direct velocity feedback the secondary force is defined to be

$$f_s(t) = -hy(t) = -h\mathbf{C}\mathbf{x}(t),$$ (4)

where h is the feedback gain, such that Eq. (2) becomes

$$\dot{\mathbf{x}}(t) = \mathbf{A}_0\mathbf{x}(t) + \mathbf{B}_p f_p(t),$$ (5)

where

$$\mathbf{A}_0 = [\mathbf{A} - h\mathbf{B}_s\mathbf{C}].$$ (6)

3. Stability

For linear systems, the stability of the system under fixed gain feedback control may be evaluated by plotting the polar response of the open loop frequency response $-hG_s(j\omega)$ in which $G_s(j\omega)$ is derived from the Fourier transform of the state equation without control

$$G_s(j\omega) = \mathbf{C}[j\omega\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}_s.$$ (7)

The Nyquist plot for the system described by the parameters given in Table 1 is shown in Fig. 3 for $h=20$. It should be noted that this Nyquist plot is only valid for the system acting within its linear range. A complete discussion of the interpretation of such a plot as the proof

Table 1. Table of model parameters.

Property	Value
m_p	0.032 kg
k_p	140 N/m
c_p	1.39 Ns/m
m_s	0.05 kg
k_s	5000 N/m
c_s	0.32 Ns/m
Force constant	1 N/A
f_0 (actuator)	10.5 Hz
f_0 (structure)	50.3 Hz
f_0 (connected system)	39.3 Hz

mass enters saturation is beyond the scope of this communication. The system is only conditionally stable in that a finite increase in gain will result in the locus encircling the $(-1,0)$ point. As the feedback gain is increased, the effective damping of the structural mass is increased, but the effective damping in the proof mass is reduced. This effect may be clearly observed in the impulse response of the system for increasing gains. Animation Mm. 1 shows the response of system to an impulse without control, $h=0$. The time series data is simulated using Matlab's ODE45 ordinary differential equation solver with variable time-steps.

Mm. 1 Animation of the system without control (287 KB). This file is of type ".mpeg".

Animation Mm. 2 shows the response of the system to the same impulse but with a velocity feedback controller with a gain of 20 N/ms^{-1}. It is clear from this animation that while the vibration of the structural mass has been significantly reduced, the vibration of the proof mass has been increased.

Mm. 2 Animation of the system with control (508 KB). This file is of type ".mpeg".

In both of the previous examples, the magnitude of the initial impulse subjected to the structural mass was relatively small such that the proof mass did not hit its end-stops and the
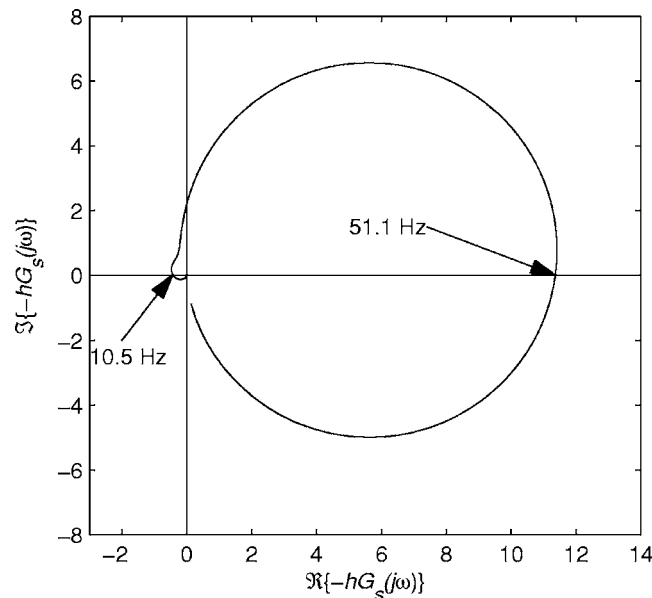
Fig. 3. Nyquist plot of the open loop frequency response function for a control gain of 20.
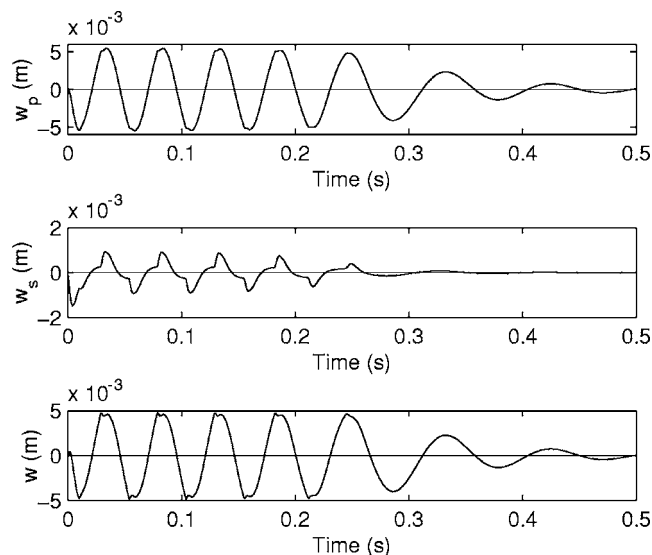
Fig. 4. Time series depicting the displacement responses of the proof and structure masses with control to a large impulse.

actuator operates within its linear range. For an actuator model with a ± 5 mm stroke and a value of $N=49$, increasing the magnitude of the primary force impulse, while maintaining a feedback gain of 20 N/ms^{-1}, causes the proof mass to saturate. Figure 4 shows the response of the system to the larger impulse. Along with the proof and structural mass displacements, the relative displacement of the two, $w(t)$, is also shown which is bounded by the imposed stroke limit of 5 mm. Again, the corresponding animation is given in Mm. 3.

Mm. 3 Animation of the system with control under saturation (768 KB). This file is of type ".mpeg".

It is clear from the figure and animation that, as the proof mass hits the end of its stroke, it imparts an impulse-like force to the structural mass but that this behavior decays away. Increasing the magnitude of the initial impulse only slightly above that used in Fig. 4 and Mm. 3 results in the responses given in Fig. 5. It is clear from this figure that the small increase in the initial force magnitude results in an unstable system. It is also worth noting that the relative displacement of the two masses at the moment of impact contains a significant amount of high frequency content, the magnitude of which increases with the velocity of the impact. The corresponding animation, Mm. 4, depicts the system becoming unstable. Note that once the system becomes unstable, after 0.7 s, control is turned off.

Mm. 4 Animation of the unstable system (1466 KB). This file is of type ".mpeg".

It can also be seen in Fig. 5 that as the system becomes unstable the oscillation frequency of both of the masses become the same and increase over time. By detecting the zero crossings of the signal we are able to plot this oscillation frequency as a function of time. This is done for both the proof and structural masses in Fig. 6. This figure shows that the oscillation of both masses starts off just above 20 Hz, close to the natural frequency of the proof mass on its suspension, but quickly increases to a value approaching 39 Hz. This corresponds to the natural frequency of the system in which the two masses are rigidly connected, oscillating on the structural support. Indeed this is the case for a significant proportion of the cycle as the proof mass is pushed up against its end-stops. As the combined structure passes through the extremes of its motion however the proof mass leaves its end-stops and, with the aid of the control force, quickly moves through its stroke, hitting the opposite end-stop, and imparting an impulse of considerable magnitude to the system, which is in phase with the velocity of the structural mass and thus acts to increase its energy.
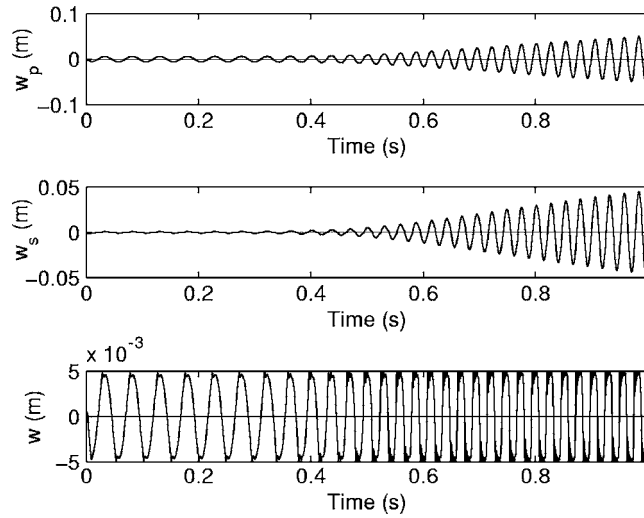
Fig. 5. Time series depicting the displacement responses of the proof and structure masses with control to a large impulse.

Attempts to overcome this by decreasing the value of N in Eq. (1), and thus softening the end-stops and increasing the time over which the impact energy is imparted to the system, has been found to make matters worse in that the initial force magnitude required to send the system unstable is reduced. Also, making the actuator damping nonlinear, in a similar way the stiffness, has been found to increase the impulse level before the instability occurs for low feedback gains, but not suppress the instability. This behavior remains a topic of ongoing research.
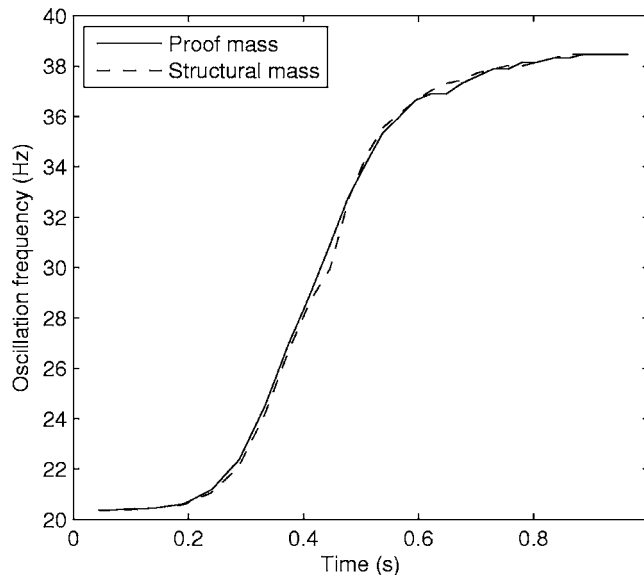
Fig. 6. Plot of the oscillation frequencies of the proof and structural masses determined from the time series of Fig. 5 as a function of time.

4. Conclusions

A model is presented of a simple structure controlled by direct velocity feedback with a saturating inertial actuator. The linear impulse response of the system shows that increasing the feedback gain of a velocity feedback controller resulted in a reduction in vibration of the structure accompanied by an increase in vibration of the proof mass. Increasing the magnitude of the initial impulse while maintaining a constant feedback gain, for which the closed loop system acting within its linear range is stable, resulted in stroke saturation and eventually instability of the feedback controller. The instability is due to the force imparted to the system as the proof mass hits its end-stops being in phase with the velocity of the structure and so reduces the damping of the system. If the initial magnitude is large enough, then the impulsive forces are large enough to overcome the damping of the system, resulting in instability.

This work presents a model of an inherently stable system with a stroke limited inertial actuator used in a velocity feedback regime. When the primary force, acting on the structure, exceeds a particular value, such that the actuator saturates, the forces due to the saturation result in the instability of the model. This agrees with the observation which has been made for practical systems in which inertial actuators are used for velocity feedback. The solution to this problem probably lies in the use of a nonlinear controller and the investigation thereof remains a topic of ongoing research.

Acknowledgments

References and links

[1] M. J. Balas, "Feedback control of flexible systems," IEEE Trans. Autom. Control **AC-23**, 673–674 (1978).

[2] S. Griffin, S. A. Lane, C. Hansen, and B. Cazzolato, "Active structural-acoustic control of a rocket fairing using proof mass actuators," J. Spacecr. Rockets **38**, 219–225 (2001).

[3] L. Benassi and S. J. Elliott, "Active vibration isolation using an inertial actuator with local displacement feedback control," J. Sound Vib. **278**, 705–724 (2004).

[4] S. J. Elliott, M. Serrand, and P. Gardonio, "Feedback stability limits for active isolation systems with reactive actuators," J. Vibr. Acoust. **123**, 250–261 (2001).

[5] C. Paulitsch, P. Gardonio, and S. J. Elliott, "Active vibration control using an inertial actuator with internal damping," J. Acoust. Soc. Am. **119**, 2131–2140 (2006).

[6] J. G. Chase, M. Yim, and A. A. Berlin, "Integrated centering control of inertially actuated systems," Control Eng. Pract. **7**, 1079–1084 (1999).

[7] D. K. Lindner, G. A. Zvonar, and D. Borojevic, "Nonlinear control of proof-mass actuator," J. Guid. Control Dyn. **20**, 464–470 (1997).

[8] D. K. Lindner, G. A. Zvonar, and D. Borojevic, "Performance and control of proof-mass actuators accounting for stroke saturation," J. Guid. Control Dyn. **17**, 1103–1108 (1994).

[9] O. N. Baumann and S. J. Elliott, "The stability of decentralized multichannel velocity feedback controllers using inertial actuators," J. Acoust. Soc. Am. **121**, 188–196 (2007).

Nonlinear frequency-dependent attenuation in sandy sediments

J. D. Holmes and W. M. Carey

*Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington St., Boston,
Massachusetts 02215*
jholmes@bbn.com, wcarey@bu.edu

S. M. Dediu and W. L. Siegmann

Mathematical Sciences Department, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180
dedius3@rpi.edu, siegmw@rpi.edu

Abstract: This paper summarizes evidence of a nonlinear frequency dependence of attenuation for compressional waves in shallow-water waveguides with sandy sediment bottoms. Sediment attenuation is found consistent with $\alpha(f) = \alpha(f_o) \cdot (f/f_o)^n$, $n \approx 1.8 \pm 0.2$ at frequencies less than 1 kHz in agreement with the theoretical expectation, ($n=2$), of Biot [J. Acoust. Soc. Am. **28**(2), 168–178, 1956]. For frequencies less than 10 kHz, the sediment layers, within meters of the water-sediment interface, appear to play a role in the attenuation that strongly depends on the power law. The accurate calculation of sound transmission in a shallow-water waveguide requires the depth-dependent sound speed, density, and frequency-dependent attenuation.

1. Introduction

The accurate calculation of sound propagation in shallow-water waveguide has been shown to depend on the use of geoacoustic profiles and the correct frequency dependence of depth-dependent attenuation profiles. Sonar engineers require a correct estimate of sound transmission loss and time spread over a wide band of frequencies. Calculations with sediment attenuation with a linear frequency dependence were found to yield inconsistent results with careful sound transmission loss and time spread measurements. However, the oceanic community has been slow in accepting the necessity for such a nonlinear dependence. Here we adopt a geophysical hypothesis that those areas primarily formed by deposition of sediments and sea level changes will have similar properties. The coast of the Eastern Gulf of Mexico, the East Coast of the United States between Cape Hatteras and Long Island, sections of the Asian Coast, the North Indian Ocean, the Strait of Korea, the East China and Yellow Seas, to mention a few, have sandy depositional layers that are classified as fast or critical-angle bottoms formed by the raising and lowering of the sea. In order to compare results between experiments consistent semantics need be used. Here the American Geophysical Union Classification system characterization of sediments composed of sand, silt, clay by particle size is employed. This characterization is shown on the diagram of Shepard[1] where such terms as sand, silt-sand, and clayey are consistent with the lower left-hand portion of this sediment triangle. The eastern coastal margins and seas examined here are between 40 and 200 m deep and have consistent bottom sediment layers formed by deposition of sands and silt.

We exclude many other coastal areas where volcanism has resulted in a thin sediment layer over layers of hard substrata. Another exception are areas determined by the depositional fans of major rivers composed of fine silts and the resulting slow bottoms, i.e., the sound speed is less than that in the overlying water. These fans are a consequence of river outflow and fine dendritic materials are constantly being added to the rivers by land runoff and human activity.

Table 1. Summary of nonlinear dependence for various sediments from previous work.

Author, date[a]	Freq. Range (Hz)	Bottom type[1,16]	C.A.	(n)	Type of Exp.[*]
Ingenito, 1973[2]	400–750	Sand	19	1.75	M
Beebe *et al.*, 1982[3]	100–600	Medium to coarse sand	29	1.76	TL
Zhou, 1985[27]	80–800	Sand-silt-clay	19	1.84	M
Zhou *et al.* 1987[8]	100–1000	Fine sand and silt	21	1.6	INV
Tatersall and Chizhik, 1992[11,12]	100–8000	Medium sand	24	2.0[**]	TL
Tappert, 1993,[10]	50–800	Sand-silt-clay	19	2.0	TL
Carey and Evans, 1998,[4]	500–1000	Sand-silt-clay	25	1.85	TL
Dediu, Carey, and Siegmann[5]					
Rozenfeld, 2001[13]	747–604	Sand-silt	23	1.8	TL
Peng *et al.*, 2004[14]	100–500	Very fine sand	21	1.65	TL
Zhou *et al.*, 2004[9]	100–700	Very fine sand	22	1.63	INV
Holmes[17,19]	200–1228	Fine sand	22	1.87	TL
			The average value is $\langle 1.8 \rangle$		$\sigma = \pm 0.2$

C.A.=Critical angle, [*]M= Modal, TL=Transmission loss, INV=Inversion. [**]Paper actually reports agreement with Biot theory, which gives $n=2.0$ in the 100–1000 Hz range.

[a]Superscript numbers refer to References and links section.

This paper summarizes experimental evidence for the exponent power, n, of the frequency dependence of the attenuation (see summary in Table 1) by the following.

$$\alpha(f) = \alpha(f_o) \cdot (f/f_o)^n, \ [\text{dB/m or Neper/m}], \quad n \approx 1.8 \pm 0.20. \quad (1)$$

The paper states that the accurate calculation of the transmission for narrowband and broadband sound in shallow-water waveguides, with a fast, sandy-silty bottom, requires non-linear frequency dependence for attenuation in the upper sediment layer for frequencies less than a few kilohertz.

2. Measurement techniques

There are four techniques for *in situ* measurement of the frequency dependence of attenuation in ocean sediments currently found in the literature.

(1) Modal techniques incorporate a vertical hydrophone array and use dispersion or wave number analysis to separate modes (e.g., see Ref. 2). Comparison of measured mode properties to theoretical properties, assuming some bottom attenuation frequency dependence, reveals the best match to data.

(2) Transmission loss (TL) techniques use TL versus range measurements at multiple frequencies with narrowband sources or broadband with impulsive sources. The estimation of frequency dependence of the attenuation is accomplished by using geoacoustic profiles that are based on geophysical measurements and forward propagation calculations. A comparison between calculated and measured TL, along with a metric to determine best fit (e.g., see Refs. 3–5), provides the exponent of the frequency dependence and a confidence interval. The constraints imposed on this process are independently measured bottom properties and, consequently, a reduction in the number of free-ranging parameters.

(3) Inversion schemes are also used and encompass many techniques, such as generic algorithms or perturbative inversions. These methods use a multi-parameter space and global search techniques, such as simulated annealing (e.g., see Ref. 6). The lack of constraints on key

parameters and knowledge of the underlying probability distributions often leads to uniqueness problems and inconsistent results with the actual bottom. These techniques are numerically intensive, have only rarely been tried over a broad frequency band, and are the subject of current research.

(4) Finally, measurement of the angle-dependent reflection coefficient successfully used at higher frequencies may also be possible with narrow beam sources and line arrays at the lower frequencies. This method often makes use of the fact that the subcritical reflection loss depends on the sediment attenuation for fast bottoms; in the case of a slow bottom, the reflection coefficient at the angle of intromission depends on the sediment attenuation (e.g., see Ref. 7).

3. Summary of attenuation measurements in shallow water

Nonlinear frequency-dependent sediment attenuation is not a new finding in shallow-water waveguides with fast, sandy bottoms. In 1973, Ingenito[2] observed a nonlinear dependence between 400 and 750 Hz. Nine years later, for two sites with medium to coarse sand, Beebe[3] found factors between 1.57 and 1.83. In the mid-1980s, Rogers *et al.*[6] summarized several previous experiments[8,9] in the Yellow Sea and observed that the frequency-dependent attenuation factor ranged between 1.6 and 1.9. In 1993, Tappert[10] found in his analysis of an experiment on the West Coast of Florida that ($n=2$) was the best-fit parameter consistent with the Biot porous medium model. Tatersal and Chizhik (1992–1993)[11,12] used the Biot theory to compute sediment parameters for a fluid bottom and an elastic bottom; calculations with a fast-field program and a multipath expansion code were found to agree with measured transmission loss over the 100 Hz to 8 kHz range when the Biot bottom was used with a frequency-dependent ($n=2$) attenuation coefficient. Additionally, Rozenfeld[13] found a frequency-dependent power of ($n=1.8$) was required to describe both narrowband and broadband TL measurements in the complex Strait of Korea. Results from experiments in the East and South China Seas conducted by the Office of Naval Research in 2003 and reported by Peng *et al.*[14] and Zhou *et al.*[9] showed dependencies between 1.6 and 1.63, while the results of Knobles *et al.*[15] were inconclusive.

Evans and Carey[4] showed that the accurate calculation of the shallow water sound transmission in a waveguide with sandy-silty bottom required a nonlinear frequency-dependent attenuation, with an exponent of ($n=1.5$), relative to a reference frequency of 50 Hz over the interval 50–1000 Hz. Dediu, Carey, and Siegmann[5] reexamined this analysis using 1 kHz as the reference frequency with the attenuation constant in the range specified by Hamilton *et al.*[16] and a refined sound speed profile determined from the experimental measurements. Extensive comparisons between measured and calculated transmission loss for frequencies between 400 Hz and 1 kHz yield a frequency power of ($n=1.85\pm0.15$) and a reference-attenuation

$$\alpha(1 \text{ kHz}, 73 \text{ m}) = 0.33 \pm 0.02, \text{ [dB/m].} \qquad (2)$$

Holmes *et al.*[17–19] conducted the Nantucket Sound Experiment to quantify the acoustic properties of the bottom sediment between 220.5 and 1228 Hz and to determine the frequency-dependent attenuation characteristics. The attenuation at 1 kHz was assumed to be consistent with Hamilton's value at that frequency. Iterative comparison of the measured to calculated transmission loss while adjusting the frequency-dependent attenuation revealed that the best fit of attenuation over all frequencies was

$$\alpha(f) = \alpha(f_o) \cdot (f/f_o)^n; \text{ with } 0.26 \leqslant \alpha(1 \text{ kHz}) \leqslant 0.27 \quad \text{and} \quad n = 1.87 \pm 0.2. \qquad (3)$$

Table 1 provides a chronological summary of previous work on the frequency dependence of attenuation, including the bottom type, the type of experiment, and the estimated critical angle of the bottom as a basis of comparison. From the results in Table 1, it is clear that there is substantial evidence that in the frequency range of 100 Hz–1 kHz, sediment attenuation follows a nonlinear frequency dependence with site dependent exponent $1.6 < n < 2.0$ for sandy-silty bottoms (see Refs. 1 and 16 for definition of sandy silt). These results compare with a summary by Zhou and Zhang[20] drawn from a larger and less restrictive group of experimental results, that is to say the inclusion of experiments with a greater uncertainty

with respect to knowledge of the bottom sediment, yielded $n=1.84$, $\alpha(f_o)=0.34$, where the reference frequency was not specified.

4. Conclusion

The accurate calculation of propagation quantities in shallow-water range-dependent waveguides with a sandy bottom boundary requires specification of the sub-bottom sound speed, density, and attenuation profiles. Normally these are based on cores, grab samples, and sub-bottom profiling. The detailed structure is usually simplified to a three- to four-layer geo-acoustic model. A key parameter in this model is the frequency-dependent attenuation profile in the near surface sediment layer. The experimental evidence summarized here shows that the near-water sediment layer has a nonlinear frequency dependence. The Biot theory[21–24] predicts an ($n=2$) frequency dependence. A consequence of the modal nature of the shallow-water waveguide is that the measured dependence should, in most instances, be less than an exponent of ($n=2$), resulting from the depth dependence of the modal functions and the geoacoustic properties of the sediment.[24] Furthermore it is known that sound speed profiles and gradients can have an important influence on the attenuation of sound; however, range-dependent calculations that use sound speed profiles in the water and geoacoustic profiles in the sediment account for these effects. For such calculations to be accurate, the use of the sub-bottom nonlinear frequency dependence is required but a dependence of ($n=1.8$) is found to provide the best fit. The ($n=1.8$) dependence has not been explained.

The purpose of this paper was to provide a summary of experimental evidence for a nonlinear frequency-dependent attenuation estimated from sound transmission measurements in shallow-water areas with sandy sediments. This nonlinear dependence is

$$\alpha(f) = \alpha(f_o) \cdot (f/f_o)^n, \quad n \approx 1.8 \pm 0.2, \quad \alpha(f_o = 1 \text{ kHz}) \approx 0.35, \text{ [dB/m]}. \qquad (4)$$

We conclude that nonlinear frequency dependence in oceanic sandy sediments is well established by the published experimental evidence.

We should also mention that early sedimentary measurements date back to those made by Wood and Weston[25] in 1964. However, their measurements considered a finer-grained low-density sediment (mud at porosity of 0.76) in Emsworth Harbor, England, and a wide frequency range of $4-50$ kHz. They demonstrated a linear dependence on frequency. In 1982, Beebe, McDaniel, and Rubano[3] determined that a linear dependence was required to provide an adequate fit to data for a mud sediment. A recent study,[26] performed south of New England in the middle Atlantic bight, for a silt-clay sediment, used a multi-parameter inversion scheme and concluded that the measurements of the attenuation values were in agreement with the previous work.[28] The physical reason for linear frequency dependence would be a decrease in the relative particle velocity of the liquid and sediment particles, as discussed by Stoll[16], pp. 1–15].

References and links

[1] F. P. Shepard, "Nomenclature based on sand-silt-clay ratios," J. Sediment. Petrol. **24**, 151–158 (1954).

[2] F. Ingenito, "Measurements of mode attenuation coefficients in shallow water," J. Acoust. Soc. Am. **53**(3), 858–863 (1973).

[3] J. H. Beebe, S. T. McDaniel, and L. A. Rubano, "Shallow-water transmission loss prediction using the Biot sediment model," J. Acoust. Soc. Am. **71**(6), 1417–1426 (1982).

[4] W. Carey and R. Evans, "Frequency dependence of sediment attenuation in two low-frequency shallow-water acoustic experimental data sets," J. Oceanic Eng. **23**(4), 439–447 (1998).

[5] S. M. Dediu, W. M. Carey, and W. L. Siegmann, "Statistical analysis of sound transmission results on the New Jersey Continental Slope," *Proceedings of the MTS/IEEE Oceans 2006 Conference*, Boston (1-424-0115 IEEE).

[6] P. Rogers, J. X. Zhou, X. Z. Zang, and F. Li, "Seabottom acoustic parameters from inversion of Yellow Sea experimental data," in *Experimental Acoustic Inversion Methods for Exploration of the Shallow Water Environment*, edited by A. Caiti, J. P. Hermand, S. M. Jesus, and M. B. Porter (Kluwer, 2000), pp. 219–234.

[7] B. Cole, "Marine sediment attenuation and ocean-bottom-reflected sound," J. Acoust. Soc. Am. **38**(2), 291–297 (1965).

[8] J. Xun Zhou, X. Zhen Zhang, P. H. Rogers, and J. Jarzynski, "Geoacoustic parameters in a stratified sea bottom from shallow-water acoustic propagation," J. Acoust. Soc. Am. **82**(6), 2068–2074 (1987).

[9] J. Zhou, X.-Z. Zhang, P. H. Rogers, J. A. Simmen, P. H. Dahl, J. Guoliang, and P. Zhaohui, "Reverberation vertical coherence and sea-bottom geoacoustic inversion in shallow water," J. Oceanic Eng. **29**(4), 988–999 (2004).

[10] F. Tappert, from correspondence (unpublished).

[11] J. M. Tatersall and D. Chizhik, "Application of Biot theory to the study of acoustic reflection from sediments," NUWC-NL Technical Report No. 10,115, Naval Undersea Warfare Center, New London, CT, September 1992; also DTIC ADA255651.

[12] J. M. Tatersall, D. Chizhik, B. F. Cole, and F. R. DiNapoli, "The effect of frequency-dependent bottom reflectivity on transmission loss in shallow water over a sandy bottom," J. Acoust. Soc. Am. **93**(4), 2395–2395 (1993).

[13] I. Rozenfeld, W. Carey, P. Cable, and W. Siegmann, "Modeling and analysis of sound transmission in the Strait of Korea," J. Oceanic Eng. **26**(4), 809–820 (2001).

[14] Z. Peng, J. Zhou, P. Dahl, and R. Zhang, "Sea-bed acoustic parameters from dispersion analysis and transmission loss in the East China Sea," J. Oceanic Eng. **29**(4), 1038–1045 (2004).

[15] D. P. Knobles, R. A. Koch, J. H. Miller, and G. R. Potty, "Evidence for nonlinear frequency dependence of attenuation in an East China Sea environment," J. Acoust. Soc. Am. **115**(5), 2551–2551 (2004).

[16] E. Hamilton *et al.*, "Acoustic and other physical properties of shallow-water sediments of San Diego," J. Acoust. Soc. Am. **28**(1), 1–15 (1956).

[17] J. D. Holmes, W. M. Carey, A. D. Pierce, and J. F. Lynch, "Attenuation characteristics of sandy sediments—a simplified Biot approach," *Proc. Int. Conf., Boundary Influences in High Frequency Shallow Water Acoustics*, Bath, England, 2005.

[18] J. D. Holmes, W. M. Carey, and J. F. Lynch, "Results from the Nantucket Sound autonomous underwater vehicle towed hydrophone array experiment," *Proceedings of the MTS/IEEE Oceans 2006 Conference*, Boston (1-424-0115 IEEE).

[19] J. D. Holmes, W. M. Carey, and J. F. Lynch, "Results from an autonomous underwater vehicle towed hydrophone array experiment in Nantucket Sound," J. Acoust. Soc. Am. Express Lett. **120**(2), EL15–21 (2006).

[20] J. X. Zhou and X. Z. Zhang, "Nonlinear frequency dependence of the effective sea bottom acoustic attenuation from low-frequency field measurements in shallow water," J. Acoust. Soc. Am. **117**, 2494 (2005).

[21] M. Biot, "Theory of propagation of elastic waves in a fluid saturated porous solid. I, low frequency range," J. Acoust. Soc. Am. **28**(2), 168–178 (1956).

[22] R. Stoll, *Sediment Acoustics* (Springer-Verlag, New York, 1989).

[23] R. Stoll, "Acoustic waves in ocean sediments," Geophysics **42**(4), 715–725 (1977).

[24] A. D. Pierce, W. M. Carey, J. F. Lynch, and M. Zampolli, "Approximate single wave equation derived from Biot's porous media equations," J. Acoust. Soc. Am. **115**(5), 2400 (2004).

[25] A. Wood and D. Weston, "The propagation of sound in mud," Acustica **14**, 156–162 (1964).

[26] G. R. Potty, J. H. Miller, and J. F. Lynch, "Inversion for sediment geoacoustic properties at the New England Bight," J. Acoust. Soc. Am. **114**(4), 1874–1887 (2003).

[27] J. xun Zhou, "Normal mode measurements and remote sensing of sea-bottom sound velocity and attenuation in shallow water," J. Acoust. Soc. Am. **78**(3), 1003–1009 (1985).

[28] D. Knobles, T. Yudichak, R. Koch, P. Cable, J. Miller, and G. Potty, "Influences on seabed acoustics in the East China Sea from distributed acoustic measurements," J. Oceanic Eng. **31**(1), 129–143 (2006).

ACOUSTICAL NEWS—USA

Elaine Moran

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

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

New Fellows of the Acoustical Society of America

Michael R. Bailey—For contributions to the therapeutic application of high intensity ultrasound.

Stephen J. Elliott—For contributions to active noise and vibration control.

Hedwig Gockel—For contributions to the understanding of pitch, intensity and modulation processing by the auditory system.

2007-08 F.V. Hunt Postdoctoral Research Fellowship awarded to Julie Oswald

The 2007–08 F. V. Hunt Postdoctoral Research Fellowship in Acoustics was awarded to Julie N. Oswald. During her Hunt Fellowship year, Dr. Oswald will undertake a research program at the Hawaii Institute of Marine Biology, University of Hawaii, Kailua, HI. The subject of her research is on an innovative approach to continuous monitoring of cetaceans in Hawaiian waters. Dr. Oswald received a B.S. degree in Marine Biology from the University of Guelph and a Ph.D. in Biological Oceanography from Scripps Institution of Oceanography, University of California, San Diego. Her Ph.D. thesis is titled "An examination of the whistling behavior of small odontocetes and the development of methods for species identification of delphinid whistles."

The Hunt Fellowship is granted each year to an ASA member who has recently received his or her doctorate or will be receiving the degree in the year in which the fellowship is to be granted. The recipient of the fellowship is that individual who, through personal qualifications and a proposed research topic, is judged to exhibit the highest potential for benefiting any aspect of the science of sound and promoting its usefulness to society. Further information about the fellowship is available from the Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Phone: 516-576-2360; fax: 516-576-2377; E-mail: asa@aip.org; Web: asa.aip.org/fellowships.html

Reviewers of Manuscripts, 2006

Each year the Journal endeavors to publish a list of all the persons who reviewed manuscripts during the preceding year. Such a list is a compendium of names supplied by the Associate Editors. Because our peer review system depends strongly on the continuing anonymity of the reviewers, the Journal publishes these names in alphabetical order without identification of the associate editors who provided the names and without identification of the papers they reviewed. The primary reason for the publication of the list is to express the Journal's gratitude to its reviewers. Reviewing a paper is often a very time consuming and demanding task, and the anonymity requirement yields no professional recognition to those who generously provide their time to help the Associate Editors decide which papers should be published and to give constructive criticisms to the authors. The Journal is justifiably proud of this list, which includes a goodly proportion of all the researchers and eminent authorities in acoustics and related fields. In a compendium of this length, omissions and errors are inevitable. If anyone notices such, please send the corrections and missing names via e-mail or regular mail to either Elaine Moran (asa@aip.org) or Allan Pierce (adp@bu.edu).

Aarts, Ronald	Akagi, Masato
Abawi, Ahmad	Akahane-Yamada, Reiko
Abramovich, Haim	Akamatsu, Tomonari
Adachi, Seiji	Akay, Adnan
Adams, Norman	Akeroyd, Michael
Adamy, Jurgen	Akhatov, Iskander
Agrawal, O. P.	Alain, Claude
Ahnert, Wolfgang	Alam, Sheikh
Ainslie, Michael	Albert, Donald
Ajaev, Vladimir	Alberts, II, W. C.

Aldridge, David
Alexander, Joshua
Alipour, Fariborz
Alku, Paavo
Allard, Jean-Francois
Allen, John
Allen, Jont
Allen, Paul
Allen, Prudence
Allyn, Hubbard
Amakasu, Kazuo
Amme, Robert
Amundin, Mats
Anastasio, Mark A.
Anderson, David
Anderson, John
Anderson, Michael
Ando, Yoichi
Andre, Michael P.
Andre-Obrecht, Regine
Anemuller, Jorn
Anfosso-Ledee, Fabienne
Apoux, Frederic
Arai, Takayuki
Aranchuk, Vycheslav
Arbisi-Kelm, Timothy
Arehart, Kathryn
Arenas, Jorge
Arieh, Yoav
Arora, Manish
Assmann, Peter
Atalla, Noureddine
Attenborough, Keith
Au, Whitlow
Auberge, Veronique
Auer, Edward
Austeng, Andreas
Aygun, Haydar
Aytekin, Murat
Backhaus, Scott
Bacon, Sid
Baddour, Ralph
Badiey, Mohsen
Badin, Pierre
Baer, Ralph
Baer, Thomas
Baggeroer, Arthur
Bailey, Michael
Baker, Richard
Baker, Steven
Balaban, Evan
Balachandran, Bala
Balasubramaniam, Krishnan
Ballagh, Keith
Bank, Balazs
Barbone, Paul
Barham, Richard
Barkmeier-Kraemer, Julie
Barlow, Claire
Barney, Anna
Barron, Michael
Barros, Allan Kardec
Barry, Johanna
Baskent, Deniz

Bass, Henry
Bastien, Francois
Bazua Duran, Carmen
Beauchamp, James
Beaujean, Pierre-Philippe
Bech, Soren
Becker, Kyle
Beddor, Patrice
Begault, Durand
Behler, Gottfried
Behne, Dawn
Beissner, Klaus
Bell-Berti, Fredericka
Benech, Nicolas
Benjamin, Kim
Benoit, Champagne
Bent, Tessa
Beranek, Leo
Berengier, Michel
Berg, Bruce
Berger, Elliott
Berman, David Harry
Bernstein Ratner, Nan
Bernstein, Joshua
Bernstein, Leslie
Bernstein, Lynne
Berry, Bernard
Berry, David
Berryman, James
Berthelot, Yves
Berthommier, Frederic
Best, Virginia
Betlehem, Terence
Beutelmann, Rainer
Bian, Lin
Bierer, Julie
Bilbao, Stefan
Bilsen, Frans
Bissinger, George
Bistafa, Sylvio
Biwa, Tetsushi
Bjor, Ole-Herman
Black, Alan
Blacklock, Oliver
Blackstock, David
Blamey, Peter
Blanc, Silvia
Blanc-Benon, Philippe
Bland, Roger
Blauert, Jens
Blottman, John
Bobrovnitskii, Yuri
Boe, Louis-Jean
Boersma, Paul
Boex, Colette
Boike, Kumiko
Boisvert, Jeffrey
Bonastre, Jean-Francois
Bond, Zinny
Bonilha, Murilo
Bonner, Brian
Boothroyd, Arthur
Bostrom, Anders
Botteldooren, Dick

Bou Matar, Olivier
Bouchage, Geraldine
Boulanger, Patrice
Boutin, Claude
Bouzidi, Youcef
Boyce, Suzanne
Braasch, Jonas
Bradley, David
Bradley, John
Braida, Louis
Brammer, Anthony
Brand, Thomas
Braun, Martin
Breazeale, Mack
Brennan, Michael
Brennan, Robert
Brentner, Ken
Bresin, Roberto
Bressmann, Tim
Bronkhorst, Adelbert
Brooks, Bennett
Browaeys, Jules
Brown, David
Brown, Judith
Brown, Michael
Browning, David
Bruce, Ian
Bruecker, Christoph
Bruneau, Michel
Brungart, Douglas
Buck, John
Bucker, Homer
Buckingham, Michael
Bucur, Voichita
Budaev, Bair
Buder, Eugene
Buell, Thomas
Bunton, Kate
Burgess, John
Burnham, Denis
Burroughs, Courtney
Busch-Vishniac, Ilene
Buske, Stefan
Buss, Emily
Busse, Lawrence J.
Butler, James P
Cable, Peter
Cai, Liang-Wu
Calamia, Paul
Caliano, Giosue
Calle, Samuel
Campbell, Murray
Cannata, J.M.
Carcaterra, Antonio
Carevic, Dragana
Carey, John
Carey, William
Cargill, G. Slade
Carin, Lawrence
Carlile, Simon
Carlson, Thomas
Carlyon, Robert
Carney, Laurel
Carstensen, Edwin

Cathignol, Dominique
Cato, Douglas
Cawley, Peter
Cazzolato, Ben
Cebrian, Juli
Cetas, Thomas C.
Chahine, Georges
Chaigne, Antoine
Chait, Maria
Chambers, David
Chambers, James
Chan, Roger
Chandler-Wilde, Simon
Chapman, David
Chapman, John
Chapman, N. Ross
Chappell, Michael
Charette, Francois
Charnotskii, Mikhail
Chatterjee, Monita
Cheatham, Mary Ann
Chen, Chi-Fang
Chen, Hongbin
Chen, Sin-Horng
Chen, Yiya
Cheney, Margaret
Cheng, Arthur
Cheng, Douglas
Chimenti, Dale
Chin, Steven
Ching, Teresa
Cho, Taehong
Choi, Jee Woong
Cholewiak, Roger
Chotiros, Nicholas
Choy, Boris
Choy, Yat-sze
Chu, Dezhang
Chunchuzov, Igor
Chung, Jing-Yau
Cicek, Ismail
Cienkowski, Kathleen
Ciocca, Valter
Clark, Robert
Clark, William
Clay, Clarence
Clemins, Patrick
Clopper, Cynthia
Coady, Jeffry
Cobbold, Richard
Cobo, Pedro
Coffeen, Robert
Colburn, H. Steven
Cole, Jennifer
Collins, Leslie
Colonius, Tim
Colosi, John
Coltman, John W.
Conant, David
Cone-Wesson, Barbara
Connaughton, Martin A.
Connine, Cynthia
Conoir, Jean-Marc
Cook, Mandy

Cooke, Martin
Coombs, Sheryl
Cooper, Nigel
Cooper, Patricia
Corbin-Lewis, Kim
Cornuelle, Bruce
Corteel, Etienne
Coussios, Constantin
Cowan, Robert S.
Cox, Trevor
Coyette, Jean-Pierre
Cranch, Geoffrey
Cranen, Bert
Craun, Matthew
Cray, Benjamin
Crocker, Malcolm
Cruickshanks, Karen
Cudina, Mirko
Culling, John
Culver, Richard
Cummings, Alan
Curra, Francesco
Curtin, Suzanne
Cuschieri, Joseph
Cutler, Anne
Cutter, George
Dacol, Dalcio
Dahl, Peter
Dai, Huangping
Daigle, Gilles
Dalenback, Bengt-Inge
D'Alessandro, Christophe
Dalmont, Jean-Pierre
Dang, Jianwu
Danilov, Sergey
Darinskii, Alexander
Darwin, Christopher
Dasgupta, Nilanjan
Dasika, Vasant
Dau, Torsten
Davenny, Ben
Davidson, Lisa
Davies, William
Davis, Anthony
Davis, Barbara
Davis, Kevin
Davis, Matt
Davy, John
Daya, El Mostapha
Dazel, Olivier
De Boer, Bart
De Bree, Hans-Elias
De Cheveigne, Alain
De Coensel, Bert
De Groot-Hedlin, Catherine
De Jong, Kenneth
De Jong, Nico
De Moustier, Christian
Deane, Grant
Declercq, Nico
Decraemer, Willem
Deecke, Volker
DeFerrari, Harry
Defrance, Jerome

Degertekin, Levent
DeJong, Kenneth
DelBalzo, Donald
Delgutte, Bertrand
Dellwo, Volker
Demany, Laurent
Demer, David
Demiroglu, Cenk
Denardo, Bruce
Deng, Li
Dent, Micheal
Deschamps, Marc
Desharnais, Francine
Desmet, Wim
Destrade, Michel
Devasia, Santosh
Dhar, Sumit
D'hooge, Jan
Diachok, Orest
DiBiase, Joseph
Diebold, John
Diehl, Randy
Dilley, Laura
Dimitriadis, Emilios
Dirckx, Joris
Di Scalea, Francesco
Djelouah, Hakim
Djikstra, Semme
Doellinger, Michael
Doherty, Karen
Doinikov, Alexander
Dolson, Mark
Donaldson, Gail
Donskoy, Dimitri
Dooling, Robert
Dorman, Michael
Dosso, Stan
Doty, Michael
Doval, Boris
Dowling, Ann
Dowling, David
Draudt, Andrew
Drennan, Ward
Dreschler, Wouter
Drullman, Rob
Drumheller, Dave
D'Spain, Gerald
Dubno, Judy
Dubois, Daniele
Duda, Richard
Duda, Timothy
Duifhuis, Hendrikus
Duraiswami, Ramani
Durgin, William
Durlach, Nathaniel
Dutta, Nader
Dye, Raymond
Ebbini, Emad
Ebenezer, D.
Eddington, Donald
Eddins, David
Edwards, Brent
Ehret, Gunter
Eisner, Frank

Ekimov, Alexander
Eldredge, Jeff
Elhilali, Mounya
Elizabeth, Olson
Ellis, Daniel
Engdahl, Bo
Engwall, Olov
Epstein, Michael
Erbe, Christine
Esipov, Igor
Espinoza-Varas, Blas
Espy-Wilson, Carol
Esward, Trevor
Evans, Richard
Everbach, Erich
Eversman, Walter
Every, Arthur
Ewert, Stephan D.
Faber, Ben
Fabry, David
Fahey, Paul
Fahnline, John
Fallat, Mark
Fant, Gunnar
Farabee, Theodore
Farassat, Feri
Farina, Angelo
Fatemi, Mostafa
Faulkner, Andrew
Fawcett, John
Feeney, M. Patrick
Feijoo, Gonzalo
Fellah, Zine el Abiddine
Feng, Albert
Ferguson, Sarah
Festen, Joost
Feth, Lawrence
Feuillon, Guy
Fidell, Sanford
Fine, Michael
Finette, Steven
Fink, Mathias
Finneran, James
Finnveden, Svante
Firszt, Jill
Fisher, Karl
Fitch, W. Tecumseh
Fitz, Kelly
Fitzgerald, Tracy
Flatté, Stanley
Flege, James
Fleischer, Guy
Flemming, Edward
Fletcher, Neville
Florentine, Mary
Fokin, Vladimir
Folegot, Thomas
Folkow, Peter
Fon, Janice
Foote, Kenneth
Forbes, Barbara
Forsythe, Stephen
Foster, F.
Fourakis, Margaritis

Fowler, Carol
Fox, Colin
Fox, Robert
Fox, Warren
Frampton, Kenneth
Franceschetti, Massimo
Francis, Alexander
Franco, Francesco
Frazer, L. Neil
Frazier, Catherine
Fred, Culick
Fredericksen, Erling
Freed, Daniel
Freitag, Lee
Freyman, Richard
Friberg, Anders
Frieda, Elaina
Frijns, J.H.M.
Frisch, Stefan
Frisk, George
Frizzell, Leon
Fu, Qian-Jie
Fuks, Iosif
Fullgrabe, Christian
Fulop, Sean
Fung, K.Y.
Funnell, W. Robert J.
Furusawa, Masahiko
Fuzessery, Zoltan
Gomes Vilda, Pedro
Gabard, Gwenael
Gade, Anders Christian
Gallun, Frederick
Gammell, Paul
Gan, Rong
Garai, Massimo
Gardner, Bill
Gardonio, Paolo
Garr, Micheal
Garrelick, Joel
Garrett, Steven
Gaskell, Gareth
Gaumond, Charles
Gaunaurd, Guillermo
Gauss, Roger
Gautier, Francois
Gavin, Joseph
Gedeon, David
Gee, Kent
Gehrig, Tobias
Geller, Drew
Genis, Vladimir
Gentry, Roger
Genuit, Klaus
George, Erwin L.J.
Gerhard, Reimund
Gerhold, Carl
Gerig, Anthony
Gerratt, Bruce
Gerstein, Edmund
Gerstoft, Peter
Ghanem, Roger
Gick, Bryan
Gifford, Rene

Giguere, Christian
Gilbert, Kenneth
Gilbert, Richard
Gilkey, Robert
Giordano, Bruno
Giordano, Nicholas J.
Glasberg, Brian
Gockel, Hedwig
Godin, Oleg
Goebl, Werner
Goldinger, Stephen
Goldsworthy, Ray
Gomes, Joao Pedro
Gomez Vilda, Pedro
Gordon, Jonathan
Gordon, Karen
Gordon, Malcolm
Gordon-Salant, Sandra
Gorga, Michael
Gorriz, J. M.
Goto, Masataka
Gottfried, Terry
Gottlieb, Hans
Gough, Colin
Gracewski, Shery
Grancharov, Volody
Grant, Ken
Grantham, D. Wesley
Green, Dale
Green, Tim
Greenberg, Steven
Greene, Jr., Charles
Griffiths, Timothy
Griffiths, Timothy
Grinchenko, Victor
Groby, Jean-Philippe
Groeschl, Martin
Gros, Laetitia
Grose, John
Grosh, Karl
Gruber, Fred
Guillot, Francois
Guinan, John
Guion, Susan
Gummer, Anthony
Guo, Ningqun
Gurbatov, Sergey
Guski, Rainer
Guyer, Robert
Haberman, Robert
Hald, Jurgen
Hall III, Joseph
Hall, Marshall
Hall, Timothy
Halle, Pierre
Hambric, Stephen A.
Hammershoi, Dorte
Handegard, Nils Olav
Hanes, Peter
Hansen, Colin
Harben, Phil
Harding, Sue
Harma, Aki
Harnsberger, James

Harold, Yurh
Harrison, Brian
Harrison, Christopher
Hartmann, William
Hartung, Klaus
Hasegawa-Johnson, Mark
Hastings, Mardi
Hatfield, Julie
Haven, Wiley
Hay, Jen
Hayashi, Takahiro
Hayek, Sabih
Hayes-Harb, Rachel
Hazan, Valerie
Healy, Eric
Heaney, Kevin
Hedrick, Mark
Hefeng, Dong
Hefner, Brian
Heil, Peter
Heimann, Dietrich
Heinz, Michael
Helfer, Karen
Hellman, Rhona
Henderson, Mark
Henderson, Paul
Henyey, Frank
Hermand, Jean-Pierre
Hermes, Dik
Herring, Gregory
Hertegard, Stellan
Hertrich, Ingo
Herzel, Hanspeter
Hess, Ursula
Heydt, Richard
Hickey, Craig
Hickling, Robert
Hillenbrand, James
Hillenbrand, Joachim
Hinders, Mark
Hirata, Yukari
Hirose, Keikichi
Hirschberg, Avraham
Hirsekorn, Martin
Hirsekorn, Sigrun
Hladky-Hennion, Anne-Christine
Hodges, Dewey
Hodgkiss, William
Hodgson, Murray
Hoekje, Peter
Hogden, John
Hohmann, Volker
Hojen, Anders
Holland, Charles
Holland, Keith
Holliday, D.
Holm, Sverre
Holmberg, Eva
Holmes, Jason
Holt, Lori
Holt, Marla
Holube, Inga
Honda, Kiyoshi
Honing, Henkjan

Honorof, Douglas
Hoole, Philip
Hornsby, Benjamin
Horoshenkov, Kirill
Horowitz, Seth
Horst, Johannes
Hosokawa, Atsushi
Hosten, Bernard
Houde, John
Houser, Dorian
Houtgast, Tammo
Houtsma, Adrianus
Hovem, Jens
Howard, David
Howarth, Thomas
Howe, Bruce
Howe, Michael
Howell, Peter
Hsu, Tom T.-J.
Hu, Yi
Huang, Lixi
Huang, Xiaoyang
Huff, Dennis
Hull, Andrew
Hume, Ken
Humes, Larry
Hunter, Eric
Hunter, Lisa
Hurley, David
Hurley, Donna
Huron, David
Hursky, Paul
Hwang, Yun-Fan
Hutchins, David
Hynynen, Kullervo
Ih, Jeong-Guon
Ihlefeld, Antje
Ihlenburg, Frank
Insana, Michael
Iourtchenko, Daniil
Isabelle, Scott
Isakson, Marcia
Iskarous, Khalil
Ito, Masashi
Iverson, Paul
Jackson, Darrell
Jackson, Philip J.
Jacobs, Laurence
Jacobsen, Finn
Jaeger, Stephen
Jaffe, Jules
James, Christopher
Jang, Seung-Ho
Janik, Vincent
Jansson, Erik
Jansson, Tomas
Jarzynski, Jacek
Jaud, Virginie
Jelinek, Milan
Jen, Philip
Jensen, Finn
Jensen, Jurgen
Jensen, Kenneth
Jerger, Susan

Jessen, Michael
Jesteadt, Walt
Jiang, Jack
Jin, Craig
Jing, Xiaodong
John, Buck
Johnson, David
Johnson, Marty
Johnson, Michael
Johnsrude, Ingrid
Johnston, Patrick
Jones, Adrian
Jones, Dennis
Jones, Jeffery
Jongman, Allard
Joseph, Phillip
Joseph, Rose
Kühner, Dietrich
Kaduchak, Gregory
Kaernbach, Christian
Kain, Alexander
Kallinger, Markus
Kang, Jian
Kapralos, Bill
Karasalo, Ilkka
Kargl, Steven
Kastak, David
Kates, James
Kato, Hiroaki
Katz, Brian
Katz, William
Keefe, Douglas
Keiffer, Richard
Keith, Stephen
Kewley-Port, Diane
Khismatullin, Damir
Khokhlova, Vera
Kidd Jr., Gerald
Kiefte, Michael
Kihlman, Tor S.
Kilfoyle, Daniel
Killion, Mead
Kim, Jin-Yeon
Kim, Yang-Hann
Kim,Yoon Young
Kimura, Masao
King, Simon
Kirby, Raymond
Kirk, Karen Iler
Kistler, Doris
Kitamura, Tatsuya
Klaeboe, Ronny
Klaseboer, Ever
Klos, Jacobt
Kluender, Keith
Knight, Richard
Knobles, David
Knutsen, Tor
Kob, Malte
Koch, Robert
Kochanski, Greg
Koelsch, Stefan
Koenig, Laura
Koeppl, Christine

Koessl, Manfred
Koike, Yoshikazu
Kolaini, Ali
Kolios, Michael
Kong, Ying-Yee
Konofagou, Elisa
Konrad-Martin, Dawn
Koopmann, Gary
Kopco, Norbert
Kozuka, Teruyuki
Krane, Michael
Krause, Jean
Kreiman, Jody
Kripfgans, Oliver
Krishna, Suresh
Krishnan, Ananthanarayan
Krumbholz, Katrin
Krutyansky, Leonid
Krylov, Victor
Kucinschi, Bogdan
Kuehn, David
Kuhl, Patricia
Kull, Robert
Kumaresan, Ramdas
Kundu, Tribikram
Kuperman, William
Kurz, Thomas
Kushwaha, Manvir
Kuttruff, K. Heinrich
Kuwada, Shigeyuki
Laback, Bernhard
Lacefield, James
Lachs, Lorin
Ladd, Robert
Lafleur, Louis
Lagree, Pierre-Yves
Lam, Yiu
Lammers, Marc
Lane, Harlan
Lang, Elmar
Langley, Robin
Larose, Eric
Larson, Charles
Larson, Gregg
Lasky, Robert
Laugier, Pascal
Launer, Stefan
Lauriks, Walter
Lavandier, Mathieu
Lavery, Andone
Lawrie, Jane
Le Bot, Alain
Lee, Haksue
Lee, Sungbok
Lee, Tan
Lee, Young
Leek, Marjorie
Legrand, Olivier
Lehman, Sean
Leibold, Lori
Leishman, Timothy
Lentz, Jennifer
LePage, Eric
LePage, Kevin

Lercher, Peter
Levi, Susannah
Levitt, Harry
Lewy, Serge
Leybaert, Jacqueline
Li, Changhui
Li, Kai Ming
Li, Pai-Chi
Li, Weichang
Li, Wen
Li, Xiaofan
Liao, Zhijie
Lieberman, Philip
Lieuwen, Tim
Lilliehorn, Tobias
Lim, Chee
Lim, Raymond
Lin, Lin
Lindblom, Bjorn E.
Lindwall, Dennis
Lingevitch, Joseph
Linton, Christopher
Lipkens, Bart
Lister, Jennifer
Litovsky, Ruth
Litvak, Leonid
Liu, Daben
Liu, Qing
Lo, Kam
Lobkis, Oleg
Long, Christopher
Lopez-Poveda, Enrique
LoPresti, Janice
Lotto, Andrew
Lotton, Pierrick
Loughlin, Patrick
Louie, John
Louisnard, Olivier
Low, Siow Yong
Lowe, Michael
Lowenstein, Joanna
Lu, Ching-Ta
Luce, Paul
Lucero, Jorge
Lucks Mendel, Lisa
Luczkovich, Joseph
Ludlow, Christy
Ludwigsen, Daniel
Lueptow, Richard
Lukashkin, Andrei
Luke, Russell
Lynch, Denis
Lynch, James
Lyon, Richard
Lyons, Anthony
Lyzenga, Johannes
Macaskill, Charlie
Mackersie, Carol
Macpherson, Ewan
Madsen, Peter
Maeda, Shinji
Maestripieri, Dario
Magnasco, Marcelo
Magnuson, James S.

Mah, Marko
Mahshie, James
Maio, Jianmin
Majdelaui, Joseph
Makarov, Sergey
Makashay, Matthew
Makov, Yuri
Makraki, George
Makris, Nicholas
Mal, Ajit
Mamou, Jonathan
Mann, David
Maradudin, Alexei
Maranda, Brian
Marengo, Edwin
Margulies, Timothy
Marshall, Bill
Marston, Philip
Martens, William
Martin, Brett
Martin, Glen
Martin, Paul
Mason, Christine
Massaro, Dominic
Mast, T. Douglas
Masters, William
Mathevon, Nicolas
Mattei, Pierre-Olivier
Mattock, Karen
Matula, Thomas
Mauermann, Manfred
Maury, Dominique
Max, Michael
Maxfield, Bruce
May, Bradford
Maye, Jessica
Mayer, Larry
Maynard, Jay
Mayo, Catherine
Mazzola, Michael
McAdams, Stephen
McCoy, John
McDaniel, James
McDonald, Bryant
McDonald, Mark
McFadden, Dennis
McFadden, Sandra
McGough, Robert
McGowan, Richard
McKay, Colette
McLaughlin, Stephen
McLeod, Sharynne
McMurray, Bob
McPherson, David
Means, Steven
Mecklenbrauker, Christoph
Meddis, Ray
Mellinger, David
Mellow, Tim
Menard, Lucie
Mercado, Eduardo
Meyer, Georg
Meyer, Jens
Michalopoulou, Zoi-Heleni

Micheyl, Christophe
Middlebrooks, John
Miksis-Olds, Jennifer
Miller, Charles
Miller, Cory
Miller, Douglas
Miller, James
Miller, James G.
Miller, James H.
Miller, Lee
Miller, Patrick
Mills, David
Milner, Ben
Minikes, Adi
Minor, Lloyd
Misun, Vojtech
Mitri, Farid
Mitson, Ron
Mitterer, Holger
Miyazaki, Kenichi
Mo, Fangshuo
Mobley, Frank
Mobley, Joel
Moffett, Mark
Mogdans, Joachim
Mohl, Bertel
Mokhtari, Parham
Molis, Michelle
Moloney, Michael
Mongeau, Luc
Montequin, Doug
Mooney, Aran
Moore, Brian
Moore, David
Moore, James
Moore, Patrick
Moorhouse, Andy
Mooshammer, Christine
Moreau, Stephane
Moretti, David
Morrison, Geoffrey
Moss, Cynthia
Moulin, Emmanuel
Mountain, David
Mourjopoulos, John
Muehleisen, Ralph
Muir, Thomas
Mullennix, John
Munhall, Kevin
Munjal, M.
Munro, Murray
Munson, Benjamin
Murphy, William
Murray, Todd
Mutlu, Erhan
Nabelek, Anna
Nachman, Adrian
Nachtigall, Paul
Nagem, Raymond
Naghshineh, Koorosh
Nagy, Peter
Naka, Yusuke
Nakai, Takayoshi
Nakatani, Tomohiro

Narayanan, Shrikanth
Naylor, Graham
Nazzi, Thierry
Neal, Steven
Nearey, Terrance
Nederveen, Cornelis
Neely, Stephen
Neff, Donna
Neild, Adrian
Nelson, David
Nelson, Doug
Nelson, Peggy
Neubauer, Juergen
Neuhoff, John
Newhall, Bruce
Newman, John
Newman, Rochelle
Nicholson, Patrick
Nielsen, Peter
Niemiec, Andrew
Nieukirk, Sharon
Niezrecki, Christopher
Nightingale, K. R.
Nilsson, Michael
Nishi, Kanae
Nittrouer, Susan
Noble, John
Noble, William
Nocke, Christian
Nolle, Alfred
Nolte, Loren
Norena, Arnaud
Norris, Andrew
Norton, Guy
Norton, Stephen
Nowacek, Douglas
Nusbaum, Howard
Nuttall, Alfred
Nyborg, Wesley
Nygaard, Lynne
Oakley, Clyde
O'Brien, Jr., William
Odom, Robert
Oelze, Michael
Oertel, Donata
Ohde, Ralph
Oldham, David
Oleson, Erin
Olofsson, Tomas
Olson, Elizabeth
Ona, Egil
Onaga, Hiroshi
Orduna-Bustamante, Felipe
Orr, Rosemary
Orris, Gregory
O'Shaughnessy, Douglas
Ostendorf, Mari
Ostrovsky, Lev
Oswald, Julie
Otani, Makoto
Ouyang, Huajiang
Ozyoruk, Yusuf
Pabon, Peter
Pace, Nicholas

Padilla, Frederic
Pagneux, Vincent
Pallier, Christopher
Palmer, Alan
Papadopoulos, Timos
Parizet, Etienne
Parker, Scott
Parks, Susan
Parra, Lucas
Parsa, Vijay
Parthasarathi, Anand
Pascal, Jean-Claude
Pasqualini, Donatella
Pastore, Richard
Patel, Aniruddh
Patel, Rupal
Patuzzi, Robert
Pavan, Gianni
Pavlovic, Chaslav
Payton, Karen
Peake, Nigel
Peat, Keith
Pecorari, Claudio
Pedersen, Michael
Pelekasis, Nikos
Pelorson, Xavier
Peperkamp, Sharon
Pereira, Jose Carlos
Perelomova, Anna
Perkell, Joseph
Perrier, Pascal
Petculescu, Andi
Petculescu, Gabriela
Peters, Robert
Peterson, Michael
Petrites, Anthony
Pfingst, Bryan
Phillips, Alana
Phillips, Dennis
Pichora-Fuller, Margaret
Picone, Joseph
Picton, Terence
Pier Paolo, Delsanto
Pierce, Allan
Pierucci, Mauro
Pishchalnikov, Yuri
Plack, Christopher
Plesniak, Michael
Poese, Matthew
Poissant, Sarah
Polka, Linda
Pollack, Gerald
Pols, Louis
Pompei, F. Joseph
Popelar, Jiri
Popelka, Gerald
Popolo, Peter
Popov, Vladimir
Port, Robert
Porter, Michael
Porter, Tyrone
Potty, Gopu
Prada, Claire-Julia
Prasanna, S. R. M.

Pratap, Rudr
Preisig, Jamesa
Premus, V.
Pressnitzer, Daniel
Preves, David
Prieve, Beth
Prosek, Robert
Prosperetti, Andrea
Pruitt, John
Purcell, David
Puria, Sunil
Qin, Michael
Qin, Qin
Qin, Yi-Xian
Quinlan, Angela
Rabbitt, Richard
Radebaugh, Ray
Radlinski, Ronald
Rafaely, Boaz
Raimbault, Manon
Rajapan, Dhilsha
Rakerd, Brad
Rakheja, Subhash
Rakowski, Andrzej
Raman, Arvind
Ramis, Jaime
Ramsay, Gordon
Rasmussen, Karsten
Rasmussen, Marianne
Raspet, Richard
Ratilal, Purnima
Ravazzani, Paolo
Ravicz, Mike
Rayess, Nassif
Raymond, Jason
Reby, David
Reed, Charlotte
Rees, Adrian
Reichmuth Kastak, Colleen
Reid, John
Remez, Robert
Ren, Tianying
Rendall, Drew
Rendell, Luke
Ressler, Patrick
Rhebergen, Koenraad
Richards, Paul
Richards, Virginia
Richardson, Mike
Richie, Carolyn
Ricketts, Todd
Riede, Klaus
Rienstra, Sjoerd
Riquimaroux, Hiroshi
Robb, Michael
Robert, Jean-luc
Roberts, Brian
Robertson, William
Robinson, Stephen
Rodman, Peter
Rogers, Catherine
Rogers, Peter
Rokhlin, Stanislav
Roman, Nicoleta

Rose, Marina
Rosen, Stuart
Rosenblum, Lawrence
Rosner, Burton
Rosowski, John
Ross, Donald
Rossignol, Clément
Rossing, Thomas
Rouas, Jean-Luc
Rountree, Rodney
Rouse, Jerry
Rouseff, Daniel
Rousseau, Martine
Roux, Philippe
Roy, Kenneth
Royston, Thomas
Rubin, Gordon
Ruch, Willibald
Rudzinsky, Jason
Rumerman, Melvyn
Rumsey, Francis
Russell, Martin
Rvachew, Susan
Ryaboy, Vyacheslav
Ryden, Nils
Sabatier, James
Sabra, Karim
Sadhal, Satwindar
Saijyou, Kenji
Sakagami, Kimihiro
Samel, Alexander
Sanchez-Morcillo, Victor
Sapozhnikov, Oleg
Sarkar, Kausik
Sarkissian, Angie
Sarradj, Ennes
Sas, Paul
Sathish, Shamachary
Savin, Eric
Sayigh, Laela
Saylor, John
Scarborough, Rebecca
Scavone, Gary
Schairer, Kim
Scharf, Bertram
Scheich, Henning
Scherer, Ronald
Schlauch, Robert
Schmidt, Henrik
Schmiedt, Richard
Schmitt, Denis
Schmitt, Douglas
Schneider, Bruce
Schomer, Paul
Schotten, Michiel
Schroeter, Juergen
Schubert, Frank
Schultz, Tanja
Scordilis, Michael
Scott, Waymond
Seeber, Bernhard
Seixas, Noah
Sek, Aleksander
Selamet, Ahmet

Sellschopp, Jurgen	Spehar, Brent	Telschow, Ken	Uppenkamp, Stefan
Serbyn, M. Roman	Spence, Charles	TenCate, James	Uscinski, Barry
Sereno, Joan	Spicer, James	Terhune, Jack	Uther, Maria
Serizawa, Hirohide	Spiesberger, John	Ternstrom, Sten	Valdivia, Nicolas
Shadle, Christine	Spoor, Philip	Tesei, Alessandra	Valeau, Vincent
Shamma, Shihab	Stafford, Kathleen	Tewary, Vinod	Valiere, Jean-Christophe
Shannon, Robert	Stanton, Timothy	Theis, Fabian J.	Valimaki, Vesa
Sharma, Anu	Starck, Jukka	Theobald, Pete	Valin, Jean-Marc
Shattuck-Hufnagel, Stephanie	Stathopoulos, Elaine	Thode, Aaron	Van Borsel, John
Shaw, Adam	Stecker, G. Christopher	Thomas, Jean-Louis	Van den Doel, Kees
Sheft, Stanley	Steele, Charles	Thomas, Rebecca	Van de Par, Steven
Sheng, Ping	Stein, George	Thomenius, Kai Erik	Van der Heijden, Marcel
Shepard, W.	Steinschneider, Mitchell	Thompson, Lonny	Van Dijk, Pim
Sheplak, Mark	Stellmack, Mark	Thompson, Stephen	Van Dommelen, Wim
Shera, Christopher	Stelmachowicz, Pat	Thomsen, Frank	Van Hemmen, J. Leo
Shield, Bridget	Stenfelt, Stefan	Thomson, Scott	Van Hoesel, Richard
Shih, Chilin	Stephanishen, Peter	Thorne, Peter	Van Lieshout, Pascal
Shinn-Cunningham, Barbara	Stevens, Kenneth	Thornton, Roger	Van Parijs, Sofie
Shokouhi, Prof.	Stevenson, Mark	Throckmorton, Chandra	Van Son, Rob
Shorter, Philip	Stickney, Ginger	Tian, Jing	Van Tasell, Dianne
Shung, K	Stinson, Michael	Tian, Yi	Van Wassenhove, Virginie
Shuvalov, Alexander	Stojanovic, Milica	Tiede, Mark	Van Wieringen, Astrid
Sibul, Leon	Stokes, Ann	Tijani, Moulay	Van Wijngaarden, Sander
Siderius, Martin	Stone, Michael	Tim, Bressmann	Varghese, Tomy
Siebein, Gary	Storey, Brian	Timmerman, Nancy	Varslot, Trond
Siegel, Jonathan	Storkel, Holly	Tindle, Christopher	Vassiliev, Dmitri
Siegmann, William	Strickland, Elizabeth	Ting, Robert	Veijola, Timo
Silverman, Harvey	Strik, Helmer	Tinianov, Brandon	Velea, Doru
Silvia, Manuel	Strube, Hans Werner	Tito, Frank	Verhey, Jesko
Simmonds, John	Studebaker, Gerald	Titze, Ingo	Verkerke, Gijsbertus
Simmons, James	Su, Zuhre	Tocci, Gregory	Verrillo, Ronald
Simonetti, Francesco	Sugimoto, Nobumasa	Toilliez, Jean	Versfeld, Niek
Simons, Dick	Sullivan, Edmund	Toivanen, Jari	Versluis, Michel
Simpson, Adrian	Sum, K. S.	Tollin, Daniel	Vieira, Jose
Simpson, Brian	Summerfield, Quentin	Tolstoy, Alexandra	Vieira, Maurilio
Simpson, Harry	Summers, Ian	Torre, Peter	Vignon, Francois
Sinex, Donald	Summers, Jason	Torres-Carrasquillo, Pedro	Vihman, Marilyn
Sivonen, Ville	Summers, Van	Torvik, Peter	Vilain, Coriandre
Skarsoulis, Emmanuel	Sun, Jie	Tourin, Arnaud	Ville, Jean-Michel
Skowronski, Mark	Sundara, Megha	Tran Van Nhieu, Michel	Vipperman, Jeffrey
Skvor, Zdenek	Sundberg, Johan	Tranter, Sue	Virovlyansky, Anatoly
Slaney, Malcolm	Supin, Alexander	Traunmuller, Hartmut	Vitevitch, Michael
Slaton, William	Surendran, Dinoj	Traykovski, Peter	Vlahopoulos, Nickolas
Slifka, Janet	Surlykke, Annemarie	Tremblay, Kelly	Vollmann, Jacqueline
Smiljanic, Rajka	Surprenant, Aimee	Tricas, Timothy	Von Hapsburg, Deborah
Smith, Bruce	Sutherland, Louis	Triep, Michael	Vorlaender, Michael
Smith, Julius	Sutin, Alexander	Trouvain, Jurgen	Voronovich, Alexander
Smith, Kevin	Suzuki, Yoiti	Trusler, J. P. Martin	Vos, Henk
Smith, Zachary	Svec, Jan	Tseng, Chiu-yu	Vos, Joos
Snell, Karen	Svensson, Peter	Tsogka, Chrysoula	Voss, Susan
Snieder, Roel	Swift, Gregory	Tucker, Robin	Waag, Robert
Snyder, Russell	Szabo, Thomas	Turgut, Altan	Wage, Kathleen
Soize, Christian	Tabain, Marija	Turicchia, Lorenzo	Wagner, Randall
Solna, Knut	Tachibana, Hideki	Turk, Alice	Wagstaff, Ronald
Sommerfeldt, Scott	Taherzadeh, Shahram	Turner, Christopher	Wahlberg, Magnus
Sommers, Mitchell	Tajima, Keiichi	Turner, Joseph	Wakefield, Gregory
Sondhi, Mohan	Talaske, Richard	Tyack, Peter	Wakeland, Ray
Song, Hee-Chun	Tam, C. K. W.	Tye-Murray, Nancy	Wales, Stephen
Southall, Brandon	Tan, Eng	Tyler, Michael	Walhberg, Magnus
Souza, Pamela	Tang, Dajun	Uchanski, Rosalie	Walker, Shane
Spahr, Anthony	Tang, S. K.	Ueda, Mitsuhiro	Wallace, Kirk
Spanos, Tim	Tarnow, Viggo	Umnova, Olga	Walsh, Edward
Sparrow, Victor	Taroudakis, Michael	Ungar, Eric	Walsh, Timothy
Spector, Alexander	Tartter, Vivien	Uppanda, Ajith	Wang, Dagen

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Acoustical News—USA   2463

Wang, Ding
Wang, Emily
Wang, Hsin-Min
Wang, Lily
Wang, Mian C.
Wang, William Shi-Yuan
Wang, Yadong
Wang, Zhisong
Wapenaar, Cornelis
Warner, Natasha
Warren, Richard
Wassink, Alicia
Waters, Kendall
Watkins, Anthony
Watson, Charles
Watson, Willie
Waxler, Roger
Wayland, Ratree
Wear, Keith
Weaver, Richard
Webb, Barbara
Weisenberger, Janet
Weismer, Gary
Wenzel, Elizabeth
Wereley, Norman
Werner, Lynne
Werner, Stefan
Westbury, John
Westwood, Evan
White, Benjamin
Whitmal, Nathaniel
Wickesberg, Robert
Wiegrebe, Lutz
Wightman, Fred
Wilcox, Paul
Wilen, Larry
Wiley, Terry
Wilhelms-Tricarico, Reiner
Williams, Earl
Williams, Kevin
Wilson, D. Keith
Wilson, Preston
Wipf, David
Withnell, Robert
Wittenberg, Thomas
Wittstock, Volker
Wochner, Mark
Wodicka, George
Wojtczak, Magdalena
Wolfe, Joe
Wolfe, Patrick
Wolfson, Michael

Wong, Kainam
Woodhouse, Jim
Worcester, Peter
Wouters, Jan
Wright, Beverly
Wright, Matthew
Wright, Richard
Wu, Junru
Wu, Kuangcheng
Wu, Sean
Xiang, Ning
Xu, Jingfeng
Xu, Li
Xu, Mubing
Xu, Yi
Yan, Zhongyu
Yang, Tsih
Yang, Wen-Bin
Yang, Xin Mai
Yao, Kung
Yevick, David
Yin, Tom
Yonak, Serdar
Yost, William
Yu, Feng
Yuen, Michelle
Yund, E.
Yunusova, Yana
Zabolotskaya, Evgenia
Zacharov, Nick
Zagrai, Andrei
Zahorian, Stephen
Zahorik, Pavel
Zajac, David
Zakharia, Manell
Zampolli, Mario
Zarzycki, Jan
Zeng, Fan-Gang
Zhang, Xiaoming
Zhang, Xuedong
Zhang, Yu
Zhang, Zhaoyan
Zhao, Yunxin
Zheng, Z. Charlie
Zhong, Pei
Zhou, Jixun
Zhu, Zhenya
Zimmer, Walter
Zorikov, Tengiz
Zsiga, Elizabeth
Zuckerwar, Allan
Zurk, Lisa

USA Meetings Calendar

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

2007

24–26 July   Revolutionary Aircraft for Quiet Communities Hampton, VA [NASA workshop hosted by the National Institute of Aerospace and cosponsored by the Joint Planning and Development Office and the Council of European Aerospace Societies. web: www.nianet.net].

2007

22–24 Oct   NOISE-CON 2007, Reno, NV [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: (515) 294-6142; Fax: (515) 294-3528; E-mail: ibo@inceusa.org].

27 Nov–2 Dec   154th Meeting of the Acoustical Society of America, New Orleans, Louisiana (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org].

2008

29 June–4 July   Acoustic 08, joint Meeting of the Acoustical Society of America (ASA), European Acoustical Association (EAA), and the Acoustical Society of France (SFA), Paris, France [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org].

27–30 July   NOISE-CON 2008, Dearborn, MI [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: (515) 294-6142; Fax: (515) 294-3528; E-mail: ibo@inceusa.org].

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

Cumulative Indexes to the Journal of the Acoustical Society of America

Ordering information: Orders must be paid by check or money order in U.S. funds drawn on a U.S. bank or by Mastercard, Visa, or American Express credit cards. Send orders to Circulation and Fulfillment Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11 per index.

Some indexes are out of print as noted below.

Volumes 1–10, **1929–1938**: JASA, and Contemporary Literature, 1937–1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5; Nonmembers $10.
Volumes 11–20, **1939–1948**: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 395. Out of Print.
Volumes 21–30, **1949–1958**: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75.
Volumes 31–35, **1959–1963**: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90.
Volumes 36–44, **1964–1968**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of Print.
Volumes 36–44, **1964–1968**: Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print.
Volumes 45–54, **1969–1973**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
Volumes 55–64, **1974–1978**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
Volumes 65–74, **1979–1983**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound).
Volumes 75–84, **1984–1988**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
Volumes 85–94, **1989–1993**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).

Volumes 95–104, **1994–1998**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632, Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound).

Volumes 105B114, **1999B2003**: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 616, Price: ASA members $50; Nonmembers $90 (paperbound).

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Acoustical News—USA   2465

ACOUSTICAL STANDARDS NEWS

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

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

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes Catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please contact Susan B. Blaeser.
Comments are welcomed on all material in Acoustical Standards News.
This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page: http://asa.aip.org.

Standards Meetings Calendar—National

Accredited Standards Committees S1—Acoustics, S3—Bioacoustics, and S12—Noise will meet in conjunction with the meeting of the Acoustical Society of America during the week of 4–8 June 2007 in Salt Lake City, UT. The U.S. TAGs to ISO/TC 43, ISO/TC 43/SC 1, and IEC/TC 29 will also meet that week.

• *Monday, 4 June 2007*

7 p.m.—ASACOS Steering Committee

• *Tuesday, 5 June 2007*

8:30 a.m.—ASA Committee on Standards (ASACOS)
10:30 a.m.—Standards Plenary Group (includes U.S. TAGs to ISO/TC 43, TC 43/SC 1, and TC 29)
2 p.m.—Accredited Standards Committee S12, Noise

• *Wednesday, 6 June 2007*

8:30 a.m.—Accredited Standards Committee S1, Acoustics
10:30 a.m.—Accredited Standards Committee, S3, Bioacoustics

Accredited Standards Committee S2—Mechanical Vibration and Shock and the U.S. TAGs to ISO/TC 108 and its subcommittees will meet in conjunction with the meeting of the Vibration Institute during the week of 18–22 June 2007 in San Antonio, TX.

• *Tuesday, 19 June 2007*

1 p.m.—Accredited Standards Committee S2, Mechanical Vibration, Shock and Condition Monitoring (includes U.S. TAGs to ISO/TC 108 and its subcommittees)

Standards News from the United States

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

American National Standards—Call for Comment on Proposals Listed

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

ASA (ASC S1) (Acoustical Society of America)

REAFFIRMATIONS

BSR S1.22-1992 (R200x) Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics [Reaffirmation of ANSI S1.22-1992 (R2002)]

Specifies standard proportions and preferred sizes of scales for plotting acoustical frequency characteristics and polar level diagrams. This Standard is not a requirement for audiograms. (12 March 2007)

BSR S1.25-1991 (R200x) Specification for Personal Noise Dosimeters [Reaffirmation of ANSI S1.25-1991 (R2002)]

Specifies performance characteristics of personal noise dosimeters that measure the percentage criterion sound exposure. It makes provision for three exchange rates: 3, 4, and 5 dB per doubling of exposure time. It provides tolerances for the entire instrument including frequency response, exponential averaging (employing SLOW and FAST), threshold, dynamic range, and other characteristics. It specifies that these tolerances be attained by the instrument in a random incidence sound field without the presence of a person wearing the instrument. (12 March 2007)

BSR S1.43-1997 (R200x) Specifications for Integrating-Averaging Sound Level Meters [Reaffirmation of ANSI S1.43-1997 (R2002)]

Describes instruments for the measurement of frequency-weighted and time-average sound pressure levels. Optionally, sound exposure levels may be measured. This Standard is consistent with the relevant requirements of ANSI S1.4-1983 (R1997), American National Standard Specification for Sound Level Meters, but specifies additional characteristics that are necessary to measure the time-average sound pressure level of steady, intermittent, fluctuating, and impulsive sounds. (12 March 2007)

REVISIONS

BSR S1.17/Part 1-200x Microphone Windscreens—Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air (Revision of ANSI S1.17/Part 1-2004)

Specifies a test to use to determine the insertion loss of windscreens for measuring microphones over a defined frequency range. The insertion loss is determined in conditions that reflect performance in still or slightly moving air. (19 March 2007)

ASA (ASC S2) (Acoustical Society of America)

REVISIONS

BSR S2.8-200x Technical Information Used for Resilient Mounting Applications [Revision of ANSI S2.8-1972 (R2006)]

Establishes the requirements to promote appropriate exchange of information regarding the application and selection of isolation for the reduction of vibrations generated by equipment and machines. Use of this standard can improve communication among engineers, manufacturers and end-users concerned with vibration isolation. (19 March 2007)

BSR S2.9-200x Parameters for Specifying Damping Properties of Materials and System Damping [Revision of ANSI S2.9-1976 (R2006)]

Presents nomenclature to improve communication among the many technological fields concerned with material damping used for resilient mountings so there will be a clear understanding by both the user and manufacturer. (19 March 2007)

IEEE (ASC C63) (Institute of Electrical and Electronics Engineers)

REVISIONS

BSR C63.19-200x Methods of Measurement of Compatibility between Wireless Communication Devices and Hearing Aids (Revision of ANSI C63.19-2006)

Applies to both wireless communications devices (WDs) and hearing aids. It sets forth uniform methods of measurement and parametric requirements for the electromagnetic and operational compatibility and accessibility of hearing aids used with wireless communications devices, including cordless, cellular, personal communications service (PCS) phones, and voice over internet protocol (VoIP) devices. This version is focused on existing services, which are in common use. (26 February 2007)

ITI (INCITS) (InterNational Committee for Information Technology Standards)

NEW NATIONAL ADOPTIONS

INCITS/ISO/IEC 14496-1-200x Information technology—Coding of audio-visual objects—Part 1: Systems (Identical national adoption and revision of INCITS/ISO/IEC 14496-1-2001, INCITS/ISO/IEC 14496-1-2001/AM1-2001, and INCITS/ISO/IEC 14496-1-2001 Amendment 4-2003)

ISO/IEC 14496 specifies a system for the communication of interactive audio-visual scenes. (12 March 2007)

INCITS/ISO/IEC 14496-2-200x Information technology—Coding of audio-visual objects—Part 2: Visual (Identical national adoption and revision of INCITS/ISO/IEC 14496-2-2001, INCITS/ISO/IEC 14496-2-2001/AM1-2002, INCITS/ISO/IEC 14496-2-2001, and INCITS/ISO/IEC 14496-2-2001/Amd 3-2003)

This part of ISO/IEC 14496 was developed in response to the growing need for a coding method that can facilitate access to visual objects in natural and synthetic moving pictures and associated natural or synthetic sound for various applications such as digital storage media, internet, various forms of wired or wireless communication, etc. (12 March 2007)

INCITS/ISO/IEC 14496-3-200x Information technology—Coding of audio-visual objects—Part 3: Audio (Identical national adoption and revision of INCITS/ISO/IEC 14496-3-2001)

ISO/IEC 14496-3 (MPEG-4 Audio) is a new kind of audio standard that integrates many different types of audio coding: natural sound with synthetic sound, low bitrate delivery with high-quality delivery, speech with music, complex soundtracks with simple ones, and traditional content with interactive and virtual-reality content. By standardizing individually sophisticated coding tools as well as a novel, flexible framework for audio synchronization, mixing, and downloaded post-production, the developers of the MPEG-4 Audio standard have created new technology for a new, interactive world of digital audio. (12 March 2007)

UL (Underwriters Laboratories, Inc.)

REVISIONS

BSR/UL 60950-1-200x Information Technology Equipment—Safety—Part 1: General Requirements (Proposal dated 2/2/07) (Revision of ANSI/UL 60950-1-2006)

Proposes changes to Annex NAD, Acoustic Tests. (4 March 2007)

Projects Withdrawn from Consideration

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

ISEA (International Safety Equipment Association)

BSR/ISEA 106-199x Hearing Protection Labeling (new standard)

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers (ASD) of the initiation and scope of activities expected to result in new or revised American National Standards (ANS). Early notification of activity intended to reaffirm or withdraw an ANS and in some instances a PINS related to a national adoption is optional. The mechanism by which such notification is given is referred to as the PINS process. For additional information, see clause 2.4 of the ANSI Essential Requirements: Due Process Requirements for American National Standards. Following is a list of proposed actions and new ANS that have been received recently from ASDs.

ASA (ASC S3) (Acoustical Society of America)

BSR S3.4-200x Procedure for the Computation of Loudness of Steady Sounds (Revision of ANSI S3.4-2005)

Specifies a procedure for calculating the loudness of steady sounds as perceived by listeners with normal hearing. The procedure is based on the spectra of the sounds. The possible sounds include simple and complex tones and bands of noise. The standard is applicable to sounds presented in free field with frontal incidence, in a diffuse field, or via headphones.

Project Need: To correct a discrepancy between the calculation described in the printed edition and that performed by the associated software. Additionally, this revision will improve the prediction of absolute thresholds as a function of frequency and equal-loudness contours. Stakeholders: Medical, research, technological developers of industrial instruments.

ASA (ASC S12) (Acoustical Society of America)

BSR S12.69-200x Railroad horn sound emission testing in locus temporis (New standard)

Provides a test method to certify the performance of warning horns installed on railroad locomotives that is not dependent on weather and other community factors.

Project Need: To develop a standard test method that permits evaluation of horn emissions consistent with 49 CFR Part 229.129. Stakeholders: Railroads, horn manufacturers, FRA, and communities surrounding railroad maintenance yards.

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.

IEEE (Institute of Electrical and Electronics Engineers)

REAFFIRMATIONS

ANSI/IEEE 656-1993 (R2006) Standard for the Measurement of Audible Noise from Overhead Transmission Lines [Reaffirmation of ANSI/IEEE 656-1993 (R2000)]: (1 February 2007)

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 the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC).

ISO Standards

ACOUSTICS (TC 43)

ISO 10848-2/Cor1:2007 Acoustics—Laboratory measurement of the flanking transmission of airborne and impact sound between adjoining rooms—Part 2: Application to light elements when the junction has a small influence—Corrigendum

ISO 140-14/Cor1:2007 Acoustics—Measurement of sound insulation in buildings and of building elements—Part 14: Guidelines for special situations in the field—Corrigendum

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 7919-3/Cor1:2007 Mechanical vibration of non-reciprocating machines—Measurements on rotating shafts and evaluation criteria—Part 3: Coupled industrial machines—Corrigendum

ISO 18431-4:2007 Mechanical vibration and shock—Signal processing—Part 4: Shock-response spectrum analysis

PLASTICS (TC 61)

ISO 6721-5/Amd1:2007 Plastics—Determination of dynamic mechanical properties—Part 5: Flexural vibration—Non-resonance method—Amendment 1

ISO 6721-6/Amd1:2007 Plastics—Determination of dynamic mechanical properties—Part 6: Shear vibration—Non-resonance method—Amendment 1

ISO 6721-7/Amd1:2007 Plastics—Determination of dynamic mechanical properties—Part 7: Torsional vibration—Non-resonance method—Amendment 1

PUMPS (TC 115)

ISO 20361:2007 Liquid pumps and pump units—Noise test code—Grades 2 and 3 of accuracy

IEC Standards

ROTATING MACHINERY (TC 2)

IEC 60034-14 Amd.1 Ed. 3.0 b:2007 Amendment 1—Rotating electrical machines—Part 14: Mechanical vibration of certain machines with shaft heights 56 mm and higher—Measurement, evaluation and limits of vibration severity

SAFETY OF MACHINERY— ELECTROTECHNICAL ASPECTS (TC 44)

IEC 61310-1 Ed. 2.0 b:2007 Safety of machinery—Indication, marking and actuation—Part 1: Requirements for visual, acoustic and tactile signals

ISO and IEC Draft Standards

25/349/FDIS ISO 80000-8 Ed.1: Quantities and units—Part 8: Acoustics (30 March 2007)

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Acoustical Standards News   2469

ADVANCED-DEGREE DISSERTATIONS IN ACOUSTICS

Grating Lobe Reduction in Transducer Arrays Through Structural Filtering of Supercritical Plates [43.38.Hz, 43.30.Yj, 43.55.Rg] —Brian E. Anderson, *Graduate Program in Acoustics, The Pennsylvania State University, State College, PA 16804, December 2006 (Ph.D).* The effect of placing a structural acoustic filter between water and the transducer elements of an array is investigated to help reduce undesirable grating lobes. A plate is mounted to transducer elements with a thin decoupling polyurethane layer between the transducer and the plate. The plate acts as a radiation/incidence angle filter to pass energy at angles near normal incidence, but suppress energy at large incidence angles. Grating lobe reduction is achieved at the expense of limiting the available steering of the main lobe. Within this steer angle limitation, the main lobe can be steered as normal while the grating lobe level is reduced by the plate's angular filtering. The insertion of a plate structural filter provides an inexpensive and easily implemented approach to extend usable frequency bandwidth with reduced level grating lobes, without increasing the number of array elements. Even though some data matches theory well, a practical material has yet to be found that possesses optimal material properties. To the author's knowledge, this work represents the first attempt to advantageously utilize a plate to provide angular dependent sound transmission filtering above the plate's critical frequency (the supercritical frequency region). [Work sponsored by ONR Code 333, Dr. David Drumheller.]

Advisor: W. Jack Hughes

Pulsed Mid-Infrared Laser Stimulation of the Auditory Nerve in the Gerbil: Implications for Cochlear Implants [43.66.Ts, 42.62.Be, 87.19.La] —Agnella D. Izzo, *Department of Biomedical Engineering, Northwestern University, Evanston, IL 60208., December 2006 (Ph.D.).* Light can artificially stimulate nerve activity *in vivo* and has the advantage of a higher spatial selectivity relative to that which can be obtained with electrical stimulation. An increased spatial selectivity of stimulation could improve the function of neuroprosthetics, such as cochlear implants, whose performance is limited by overlapping electric fields from neighboring electrodes. The studies reported here investigated optical stimulation of nerves in the gerbil auditory system. Pulsed, infrared light was used to elicit compound action potentials (CAPs) from the gerbil cochlea *in vivo*. Penetration depth was varied by changing the wavelength of irradiation, to select the tissue depth of stimulation. Experiments showed no immediate damage to the cochlea when optically stimulating at 400 Hz for several hours. Immunohistochemical staining for the protein c-FOS revealed spatial specificity of the optically stimulated cochleae: only the spiral ganglion cells directly in the optical path revealed c-FOS staining, indicating that they were stimulated. In contrast, c-FOS staining of electrically stimulated cochleae demonstrated electric current spread. An electrophysiologic measurement of the spatial specificity of optical stimulation was conducted with tone-on-light masking experiments. Results from tone-on-light masking studies revealed tuning curves that were similar in extent to tone-on-tone tuning curves.

Advisors: Joseph T. Walsh, Jr., Claus-Peter Richter

Microstructure and Acoustical Macro-Behavior: Approach by Reconstruction of a Representative Elementary Cell [43.55.Ev] —Camille Perrot, *Department of Mechanical Engineering, Université de Sherbrooke, Québec, Canada, December 2006 (Ph.D.).* The fundamental issue of determining acoustic properties of porous media from their local geometry is examined in this Ph.D. dissertation thesis, thanks to a sample of open-cell aluminium foam analyzed by axial computed microtomography. Various geometric properties are measured to characterize the experimental sample at the cell size level. This is done in order to reconstruct a porous medium by means of idealized three- and two-dimensional unit cells. The frequency dependent thermal and velocity fields governing the propagation and dissipation of acoustic waves through rigid porous media are computed by Brownian motion simulation and the finite element method, respectively. Macroscopic behavior is derived by spatial averaging of the local fields. Our results are compared to experimental data obtained from impedance tube measurements. First, this approach leads to the identification of the macroscopic parameters involved in Pride and Lafarge semiphenomenological models. Second, it yields a direct access to thermal and viscous dynamic permeabilities. However, the bidimensional model underestimates the static viscous permeability as well as the viscous characteristic length; what thus require a three-dimensional implementation.

Advisors: Raymond Panneton and Xavier Only

BOOK REVIEWS

P. L. Marston

Physics Department, Washington State University, Pullman, Washington 99164

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

Multiple Scattering, Interaction of Time-Harmonic Waves with N Obstacles

P. A. Martin

Cambridge University Press, 2006. 450 pp. Price: $140.00 (hardcover). ISBN: 0-521-86554-9

In many fields such as acoustics, elastodynamics, electromagnetism or hydrodynamics, characterizing the interaction of a wave-field with two or many obstacles is an important issue. In case of a large number of obstacles, they may be placed at deterministic, periodic or random positions. This is what "Multiple Scattering" is all about. This problem, which takes its roots in the 19th century, and has been in constant evolution during the 20th century, is still up to date. No doubt the next to come improvements in computing devices will lead to its further development in the 21st century. The adventure of "Multiple Scattering" is far from being over…

Let's state it clearly, no one interested in multiple scattering can afford to ignore Paul Martin's book. Reading it will gain precious time for the beginners and will sure teach a lot to the others. The topic is, of course, too wide to be covered in all details. P. Martin invites us to a wonderful journey from the 19th century to the present time, through the fascinating reading of his paragraphs entitled "Literature," with short stops to investigate in more detail selected specific aspects of Multiple Scattering. Chapters 2 and 3, for example, could be summarized as "*Every thing you always wanted to know about 'addition theorems,' but were afraid to ask*" (W.A. Key, 1972). The book contains a sprinkling of quotations from the literature, which are not humorless.

Methods based on separation of variables are extensively covered in Chapter 4. They apply more specifically to simply shaped obstacles (cylinders, spheres,…). In my opinion, simply shaped obstacles are very interesting and will remain so for still a long time. First, many manufactured objects are indeed simply shaped, or result from the assembling of such objects. Second, for large numbers of randomly placed obstacles, one has to consider mean fields. Calculating the exact scattered field from each single object thus becomes less important. Whenever the shape of an object is complicated, a simpler one may often fruitfully approximate it.

Chapters 5 and 6 are devoted to Integral Equation Methods. These methods are not specific of multiple scattering and both chapters will certainly attract a large audience. Not willing to imitate P. Martin, I would say that these chapters contain many results "that are well known to those to whom they are well known." In particular, the Fast Multipole Method is currently a thriving research area. Here, P. Martin tries to emphasize those features that are characteristic of multiple scattering. The very popular *T*-matrix method and the null-field method are developed in Chapter 7. Those three chapters are written with a remarkable sense of synthesis.

The last chapter is dedicated to approximations. That term is to be understood in a very large meaning, as one can find there an introduction to a wide research area including effective medium theories such as Foldy's, Lax's, Twersky's, Waterman and Truell's, Fikioris and Waterman's, Llyod and Berry's, P. Martin's, and many others.

P. Martin's big picture of Multiple Scattering is exceptional and his book shows the coherency of the methods, beyond "technical details," so that multiple scattering appears now as an integral part of mathematical physics. This book is somewhat atypical and has had a very long gestation period. It contains many comments on the origin of the methods, their application fields, their strengths and their limits, and all that with citations from the best authors. This gives the reader an overview that will help him in the choice of the method appropriate to his problem. Such an approach, while relatively rare and difficult to keep to the end, is a worth noticing conducting line throughout the book. P. Martin's book is undoubtedly a most precious guide in the extensive field of multiple scattering.

I like everything in P. Martin's book, even the appendices, in particular appendix D (One-dimensional finite-part integrals). Personally, I think this book rehabilitates Twersky's formidable work in multiple scattering. As nobody is perfect, it certainly misses a few references to phononic crystals, for which the Multipole Method has been playing an important role in the study of the influence of "defects" in crystals. I have forgotten to say that P. Martin's style is concise, clear, and shows the author's will to always point directly to the essential and not leave place to pointless comments. I shall conclude with a quotation: "*ce que l'on conçoit bien s'énonce clairement et les mots pour le dire arrivent aisément*" (N. Boileau, l'art poétique 1674, or P. Martin, Multiple Scattering 2006). I hope P. Martin's book will receive all the success it deserves.

JEAN-MARC CONOIR
LAUE, Université du Havre
Place R.Schuman
76610 Le Havre
France

How Equal Temperament Ruined Harmony (and Why You Should Care)

Ross W. Duffin

W. W. Norton, 2006. 196 pp. Price: $25.95 (hardcover). ISBN-13: 978-0-393-06227-4

Analysis, Synthesis, and Perception of Musical Sounds

James W. Beauchamp, Editor

Springer, 2007. 325 pp. Price: $89.95 (hardcover) ISBN-13: 978-0-387-32496-8

Here are two very different books, which serve as an interesting illustration of the broad range of ideas we must consider to advance both our understanding of how we hear music and our ability to produce it effectively. Each in its own way challenges us to reconstruct how we hear and how we perform.

Ross Duffin is a teacher of music history and performance practice, specializing in early music. His book title might suggest a mere rant, but one should not be put off by this as I was at first. Yes, it is unapologetically polemical, but cheerfully so, as well as amusingly and rationally. We get a knowledgeable historical survey of a topic that deserves more attention than it often gets, and careful consideration of how modern performers ought to make use of this knowledge. Short biographical sketches account for 16% of the text, adding human interest and helping avoid unrelieved abstract tuning theory and musicological argument.

There are real conflicts to be faced, and real compromises to be made. Shall we ask for melodic expressivity, and especially for leading tones which edge closer to their intended resolution? Or shall we place higher value on pure tuning of thirds? One leads to G sharps that are higher than A flats and the other to the exact opposite. Equal temperament (ET) throws up its hands in surrender, rather than struggle for truly artful solutions to this problem.

It is no surprise to read that J. S. Bach's *Well-Tempered Clavier* was intended for "circulating" unequal temperament, as the argument that Bach

was advocating equal temperament has always been weak. More important, while the default "modern" attitude has been that equal temperament had already won the field early in the 19th century, Duffin makes a case that unequal temperaments continued to have an important presence even up to the beginning of the 20th. One of his prime examples is a conflict between schools of string playing, exemplified by Joseph Joachim and Pablo Sarasate, who concretely illustrated the tendency to depart from ET in opposite directions! He even cleverly turns on its head the oft-quoted advice of Louis Spohr in 1832 ("By pure intonation is naturally meant that of equal temperament, since in modern music no other exists"), arguing that Spohr would not have been so vehement if he were not fighting against an alternative that was alive and well. Spohr may instead have been recommending ET mainly for inexperienced beginners, who were not yet ready to deal with more artful nuances.

Those artful nuances prove a formidable challenge to modern technology in the volume edited by James Beauchamp, which is the latest addition to Springer's series on Modern Acoustics and Signal Processing. Forty years ago it must have seemed that it would be straightforward to analyze and re-synthesize sounds with Fourier methods and increasingly powerful computers, but many subtle details have provided continuing challenges to our ingenuity. Professor Beauchamp has been involved throughout these developments, and is well qualified to bring us this interesting symposium.

One gets a good historical introduction and overview from Beauchamp's Preface as well as the first section of Chapter 4. From the first roots of the vocoder in the 1930s, we can see the methods become much more powerful with the Fast Fourier Transform in the 1970s. More recently, fast personal computers have opened up many practical applications for mass consumption. Clearly many of the same concepts and techniques have par-allel developments in speech coding, but the focus in this volume is on applications to music.

Individual chapters begin with Beauchamp describing his work (with various collaborators) on basic synthesis of musical instrumental sounds. This is followed by Judith Brown's work on pitch tracking, and explorations of timbre morphing by Lippold Haken, Kelly Fitz, and Paul Christensen. Chapter 4 presents the important work of Scott Levine and Julius Smith on efficient modeling of general musical sounds as the sum of sines plus transients plus noise. Xavier Rodet and Diemo Schwarz describe work on estimation and representation of spectral envelopes, and in Chapter 6 Andrew Horner discusses FM synthesis and its relation to other methods. The final two chapters, by John Hajda and by Sophie Donnadieu, explore in depth the perception and representation of timbre.

All chapters include extensive references for those who wish to pursue additional details. The publisher, editor, and authors are all to be commended for a uniformly good job of presentation, including graphs, typography, and readable prose.

One final illustration of how the compartmentalization of our specialties leaves us with further delicate tasks to undertake: Judith Brown's work on musical pitch tracking was kept relatively simple by assuming a twelve-tone equal-tempered scale (p. 93). If we want to accomplish the same task in the environment of unequal temperaments advocated by Duffin, we must issue an invitation for some clever extension of Brown's work!

DONALD E. HALL
California State University
Sacramento CA 95819-6084

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
ALIREZA DIBAZAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DIMITRI DONSKOY, *Stevens Institute of Technology, Castle Point on the Hudson, Hoboken, New Jersey 07030*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden 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, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

7,134,341

43.35.Zc METHODS AND DEVICES FOR DETERMINING THE RESONANCE FREQUENCY OF PASSIVE MECHANICAL RESONATORS

Doron Girmonsky *et al.*, assignors to Zuli Holdings Limited
14 November 2006 (Class 73/579); filed 21 April 2004

To measure pressure within a blood vessel or cranial cavity, a pressure sensor **20** is claimed where rigid Pyrex substrate **22** carries plate **24** with a number of millimeter-wide cavities **28** (with dimension **L1**), each with a 25-micron thick membrane cap **29**, which will resonate at about 60 kHz. The membrane's resonant frequency is altered by the pressure difference between the fluid environment pressure **20** (to be measured) and that of the
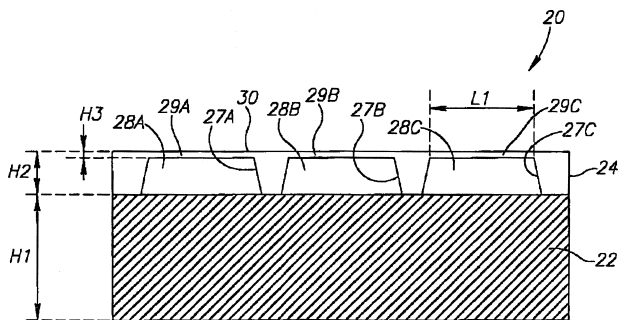
cavity volume **28**. An ultrasound interrogation beam contains wide-band energy in the ~60 kHz range and a 3 MHz carrier. Back-scattered 3 MHz carrier sound energy includes Doppler-shifted frequency components according to the excited vibration of the membranes of various cavities. Calibration is predetermined by tests and recorded in a lookup table for future data analysis of fluid pressure results.—AJC

7,141,917

43.35.Zc INDIUM OR TIN BONDED ACOUSTIC TRANSDUCER SYSTEMS

Mark J. Beck and Richard B. Vennerbeck, assignors to Product Systems Incorporated
28 November 2006 (Class 310/334); filed 10 June 2005

This is an assembly method for ultrasonic cleaners that operate at a frequency between 0.4 and 3 MHz. Piezoelectric crystals **130** are attached to
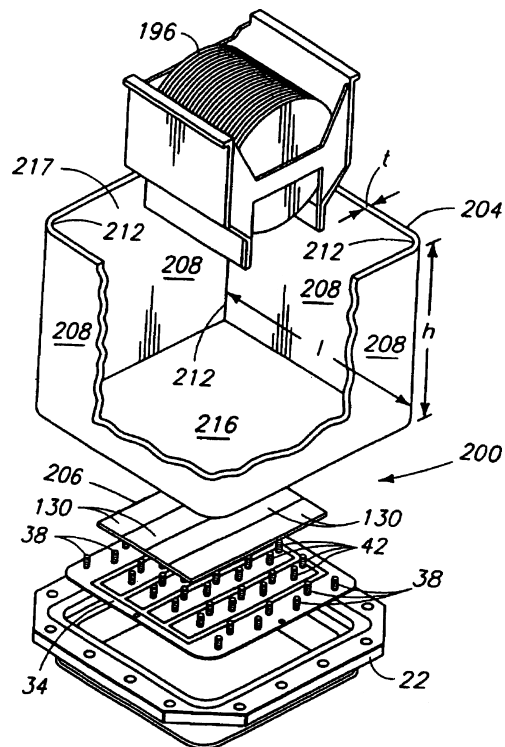
the bottom **216** of a fluid reservoir **212**. Contacts **42** supply electrical excitation. The patent states that indium or tin solder of at least 99.99% purity be used to attach the transducers to the reservoir bottom.—AJC

7,134,319

43.38.Ar ACOUSTIC WAVE SENSOR WITH REDUCED CONDENSATION AND RECOVERY TIME

James Z. Liu, assignor to Honeywell International Incorporated
14 November 2006 (Class 73/31.06); filed 12 August 2004

This author discloses a sensor that really seems more like a hodge-podge of sensors and actuators thrown together in the hopes that the deficiencies of one will be ameliorated by some combination of the others. Heaters, absorbent materials, SAW oscillators, and all other acoustic modes are invoked to sort out the signals that come from the condensation of water on the surface of a passivated substrate. One can only guess at how this all works out after the signal processing is applied, so there is really not much to learn from this patent.—JAH

7,136,494

43.38.Ar VEHICLE ACCESSORY MICROPHONE ASSEMBLY HAVING A WINDSCREEN WITH HYDROPHOBIC PROPERTIES

Alan R. Watson *et al.*, assignors to Gentex Corporation
14 November 2006 (Class 381/86); filed 6 August 2003

This extensive patent deals with microphones intended for hands-free application in an automobile. Multiple microphones may be combined in, say, a rear-view mirror structure, and their outputs combined via downstream signal processing, including gating and adaptive equalization, in or-

der to enhance the speech-to-noise ratio. A number of structural aspects are described as well. This is recommended reading for designers of systems in this general category.—JME
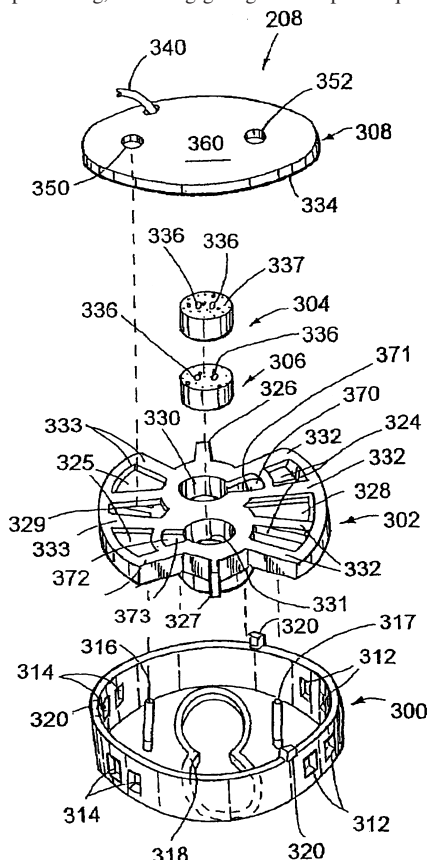
7,136,496

43.38.Ar ELECTRET ASSEMBLY FOR A MICROPHONE HAVING A BACKPLATE WITH IMPROVED CHARGE STABILITY

Aart Z. van Halteren *et al.*, assignors to Sonion Nederland B.V.
14 November 2006 (Class 381/174); filed 8 October 2002

This patent addresses methods of protecting the electret layer on the microphone's backplate from cumulative degradation due to dirt and moisture. Such measures may be especially needed for microphones used in hearing aids and other applications of wide environmental exposure.—JME

7,136,497

43.38.Ar ACOUSTICAL SWITCH FOR A DIRECTIONAL MICROPHONE

John P. McSwiggen, assignor to Knowles Electronics, LLC.
14 November 2006 (Class 381/313); filed 17 April 2003

This patent deals with the design of a mechano-acoustic switch, easily accessed by the wearer of a hearing aid, for shifting the response from omnidirectional to unidirectional as may be required by acoustical environmental conditions.—JME

7,136,500

43.38.Ar ELECTRET CONDENSER MICROPHONE

James S. Collins, assignor to Knowles Electronics, LLC.
14 November 2006 (Class 381/369); filed 5 August 2003

The patent deals with the design and construction of small gradient-type electret microphones. The specific goals here are structural simplicity

2476   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Reviews of Acoustical Patents   2476
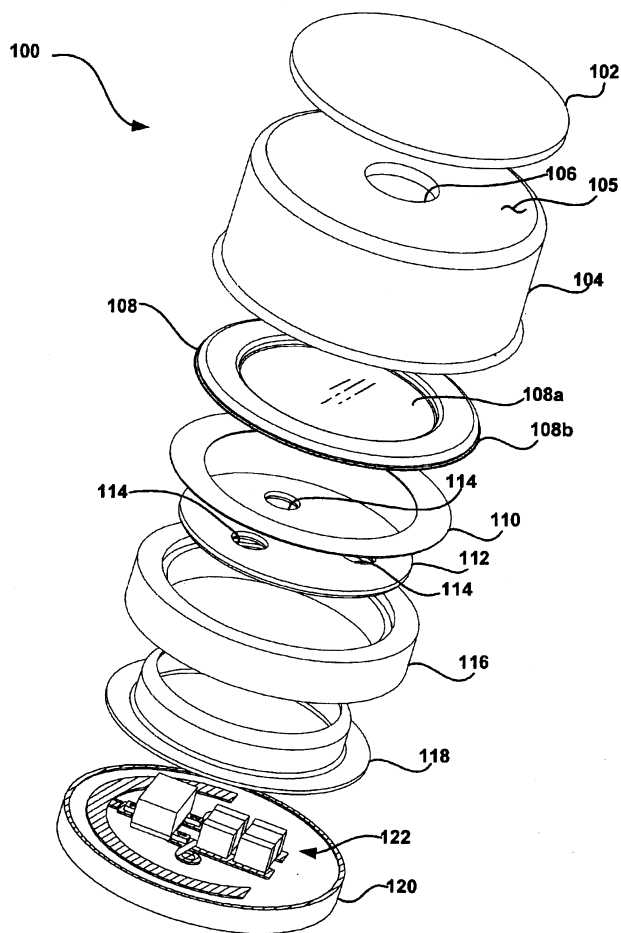
and the maintenance of uniform directional response through the use of acoustical resistances whose characteristics are both uniform and stable over long periods of time.—JME

7,142,684

43.38.Ar MICROPHONE ASSEMBLY

Megumi Horiuchi and Tsutomu Ojima, assignors to Citizen Electronics Company, Limited
28 November 2006 (Class 381/361); filed in Japan 17 May 2001

The patent basically discusses design and assembly techniques for producing small electret microphones with their associated circuitry. It would be of interest to those who manufacture cellular phones and other high-volume miniature electronic products.—JME

7,140,250

43.38.Bs MEMS TEETER-TOTTER ACCELEROMETER HAVING REDUCED NON-LINEARTY

Ronald B. Leonardson and David L. Malametz, assignors to Honeywell International Incorporated
28 November 2006 (Class 73/504.14); filed 6 June 2005

This patent discloses the design of interleaved capacitance electrodes to linearize the capacitance vs. displacement output of a parallel-plate capacitor. Nothing new here.—JAH

7,103,190

43.38.Ct ACOUSTIC DEVICE

David Johnson and William John Metheringham, assignors to Newlands Technology Limited
5 September 2006 (Class 381/152); filed in United Kingdom 21 August 2001

Giant magnetostrictive transducer **5** is affixed to rod **4**, which is, in turn, affixed via loop **2** at lower edge **3** to flexible membrane **1**. When an
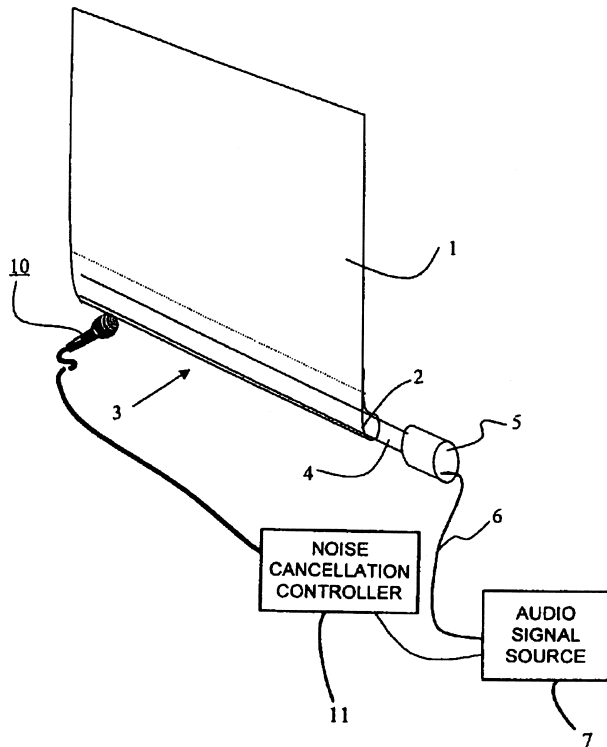
audio electrical signal is applied to **5**, membrane **1** is excited by rod **4** and thus becomes a loudspeaker. Microphone **10** and module **11** can be used to generate an antiphase signal whereby the device is said to be able to reduce perceived noise in a room.—NAS

7,103,195

43.38.Dv SPEAKER UNIT

Tomohiro Kawata and Kunio Mitobe, assignors to Pioneer Corporation
5 September 2006 (Class 381/412); filed in Japan 17 September 1996

Although the voice coil is circular and the diaphragm is described as elliptical, this speaker features a top plate **1**, magnet **2**, back plate and yoke
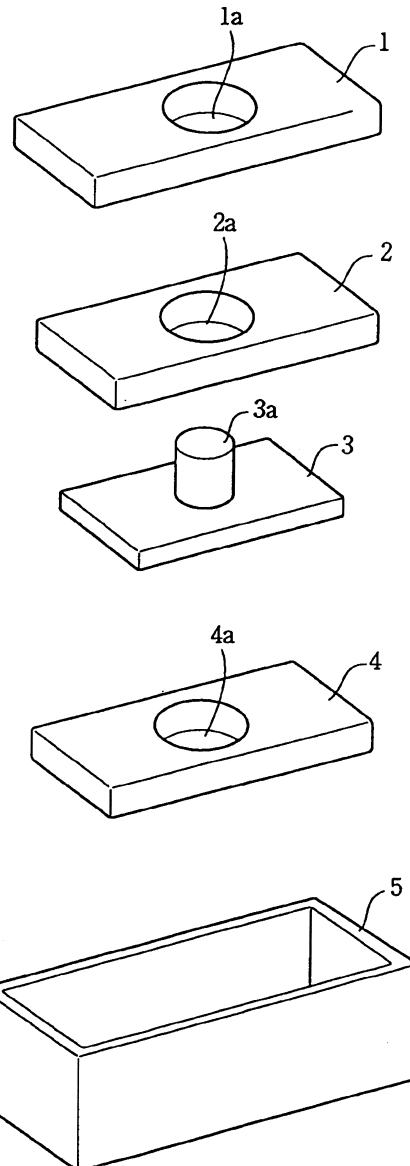
3, magnet **4**, and case **5** that are rectangular. The device is designed to be used in TV sets.—NAS

7,110,561

43.38.Dv TRANSPARENT PANEL-FORM LOUDSPEAKER

Tai-Yan Kam, assignor to Neosonica Technologies, Incorporated
19 September 2006 (Class 381/152); filed in Taiwan 28 December 2000

A discussion of what is said to be the theoretical basis for the invention precedes a description of the preferred embodiments of a transparent panel-form loudspeaker **10**, which can consist of many clear materials (including PS, PVC, PMMA, PET, and PC) and which can use various and sundry
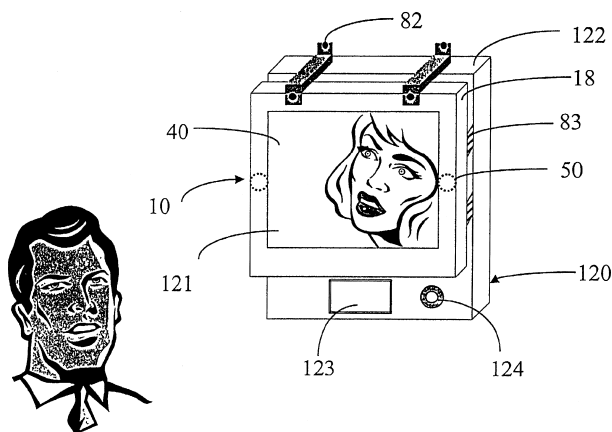
motor topologies with applications ranging from front projection screens, cellular telephones, video intercom (see figure), video cameras, and personal digital assistants.—NAS

7,110,565

43.38.Dv ELECTROACOUSTIC TRANSDUCER WITH A DIAPHRAGM, AND METHOD FOR FIXING A DIAPHRAGM IN SUCH TRANSDUCER

Wilmink Engbert *et al.*, assignors to Sonionmicrotonic Nederland B.V.
19 September 2006 (Class 381/418); filed in Netherlands 6 April 1999

A means of attaching diaphragm **6** to the case **3** using a polymer film,
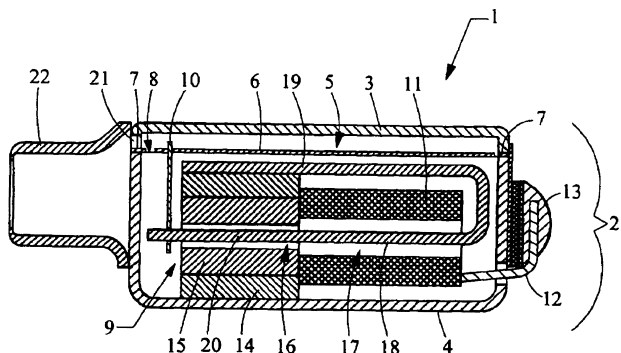
which also is said to improve the manufacturing of a balanced armature hearing aid transducer, is described in terse prose and 20 claims.—NAS

7,132,597

43.38.Dv TRANSDUCER FOR CONVERTING BETWEEN MECHANICAL VIBRATION AND ELECTRICAL SIGNAL

David Lee Hosler, assignor to Taylor-Listug, Incorporated
7 November 2006 (Class 84/723); filed 26 February 2002

This transducer translates guitar body-vibration bending modes into an audio signal. Sensor **500** is mounted on guitar interior surface at **509**. Magnet **510** is supported on a spiral spring that translates body twisting motion
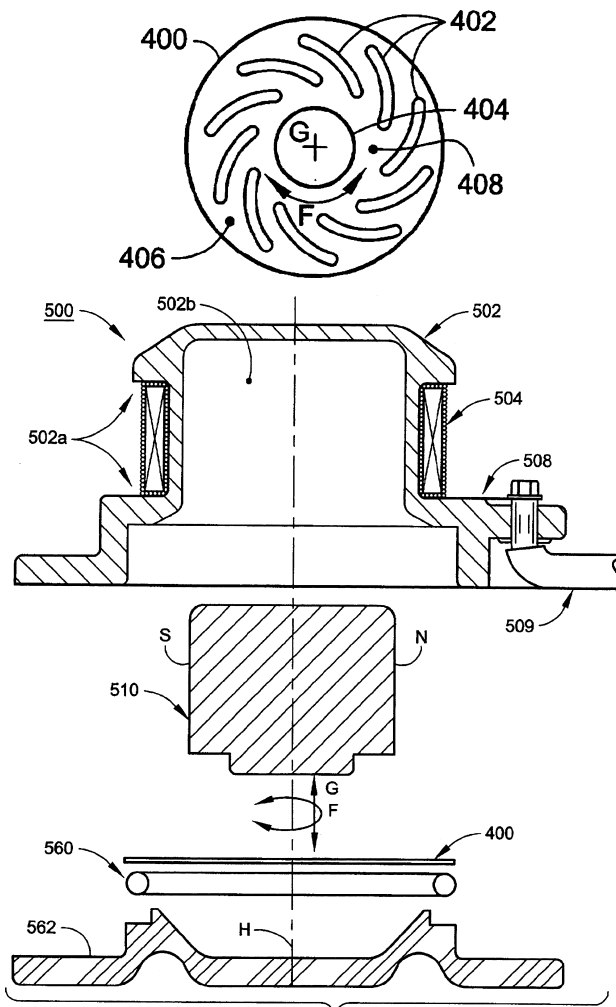
F into axial motion **G** along axis **H**. Motion of magnet **510** on axis **H** while within coil **504** produces a signal current which can be amplified to produce sound with a unique timbre.—AJC

7,136,495

43.38.Ew LOUDSPEAKER DEVICE

Hidekazu Tanaka and Fumiyasu Konno, assignors to Matsushita Electric Industrial Company, Limited
14 November 2006 (Class 381/96); filed in Japan 19 May 1999

In a large-screen television set, the only practical location for loudspeakers may be somewhere behind the screen, with sound conducted from
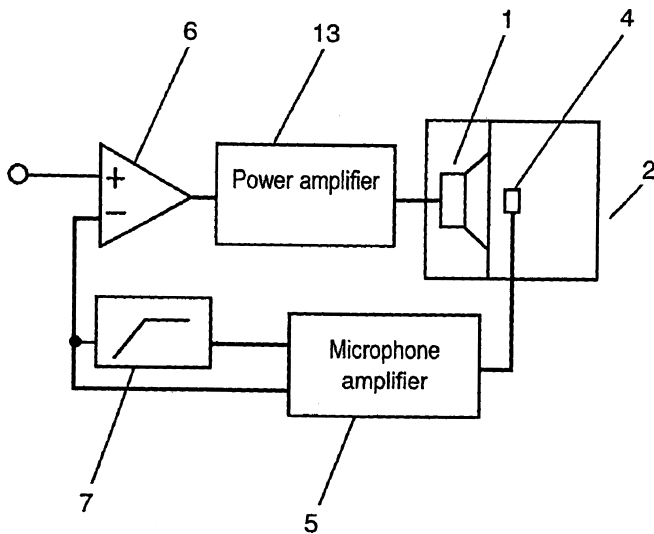
the speakers to the front of the cabinet through elongated waveguides. This interesting patent teaches that pipe resonances can be controlled by a filtered active-feedback system.—GLA

7,129,799

43.38.Fx PIEZOELECTRIC VIBRATION ELEMENT AND PIEZOELECTRIC FILTER

Yukinori Sasaki, assignor to Matsushita Electric Industrial Company, Limited
31 October 2006 (Class 331/158); filed in Japan 2 April 2001

This patent describes the various uses of ZnO as a piezoelectric material for creating bulk acoustic wave oscillators and filters. The stated objective is to disclose how to get reasonably high Qs and also confine the resonance behavior to the fundamental mode. There is satisfying detail about the different modes (e.g., thickness, transverse) that may be used, and conductor configurations to excite them. The ideas are not new, but the application to ZnO may be.—JAH

7,138,750

43.38.Fx PIEZOELECTRIC GENERATOR FOR SENSORS INSIDE VEHICLE TIRES

Federico Mancosu *et al.*, assignors to Pirelli Pneumatic S.p.A.
21 November 2006 (Class 310/339); filed in the European Patent Office 18 April 2000

This patent discloses the use of piezoelectric coaxial cable to harvest power from a rolling tire. This type of cable, typically used for intrusion detection and perimeter monitoring, is shown to generate significant
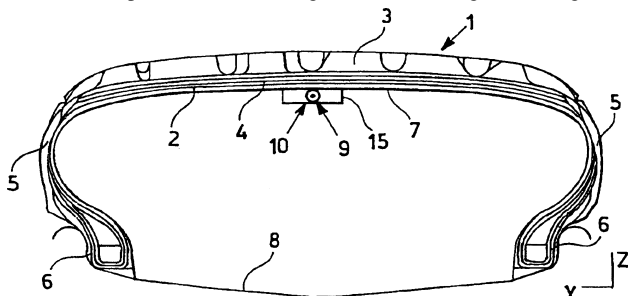
amounts of power when applied to the inside of a tire. The patent provides supporting data and some alternate configurations for the cable.—JAH

7,138,889

43.38.Fx SINGLE-PORT MULTI-RESONATOR ACOUSTIC RESONATOR DEVICE

Kenneth Meade Lakin, assignor to TriQuint Semiconductor, Incorporated
21 November 2006 (Class 333/189); filed 22 March 2005

This patent discloses the design of a vertically stacked film bulk acoustic wave resonator that has separate transducer and resonator parts. Although this may have been new when the author first conceived it, there have been other patents covering the same concept published during the last few years.—JAH

7,133,530

43.38.Hz MICROPHONE ARRAYS FOR HIGH RESOLUTION SOUND FIELD RECORDING

Mark Alistair Poletti, assignor to Industrial Research Limited
7 November 2006 (Class 381/92); filed in New Zealand 2 February 2000

The patent describes a microphone array with 8 cardioid elements positioned at equal angles in the horizontal (azimuthal) plane. The intent is to pick up a "surround sound" field in a typical recording environment via resolution of the outputs of the individual microphones into "spatial harmonics." These are then combined into first, second, or even higher order
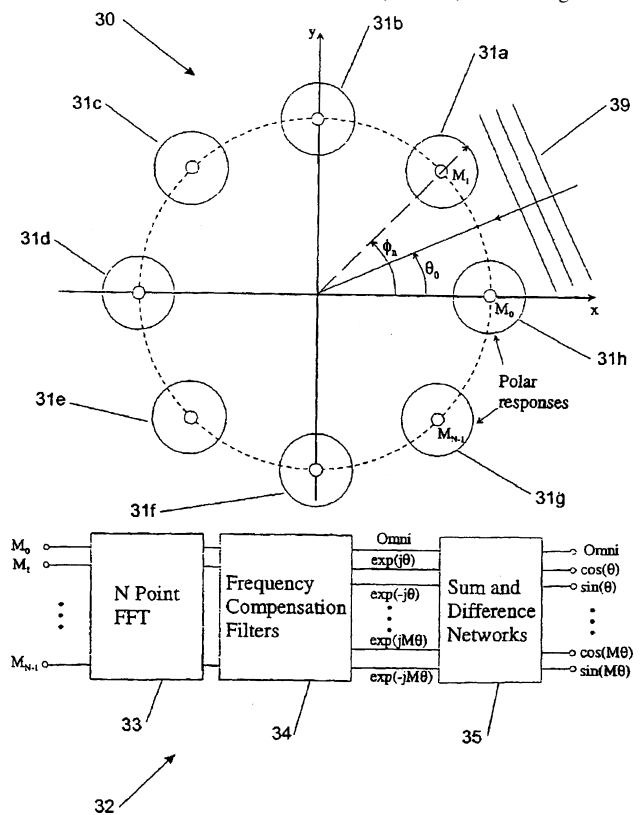
patterns, to enhance the ultimate sonic resolution (acutance) of the array. This is all accomplished through linear signal combination, inducing equalization and delay elements. There is a vast body of prior art in this area, dating back to Michael Gerzon's development of the Sound Field microphone in the mid-1970s.—JME

7,142,677

43.38.Hz DIRECTIONAL SOUND ACQUISITION

Aleksandr L. Gonopolskiy and Gamze Erten, assignors to Clarity Technologies, Incorporated
28 November 2006 (Class 381/92); filed 17 July 2001

We can all understand that the minor (back) lobe of a hypercardioid microphone exhibits higher directionality than the front lobe—but exhibits less sensitivity. What if we could turn the microphone around; take advantage of the sharper back lobe; and somehow suppress the major lobe? The authors of this patent suggest that arrays of gradient microphone pairs, with
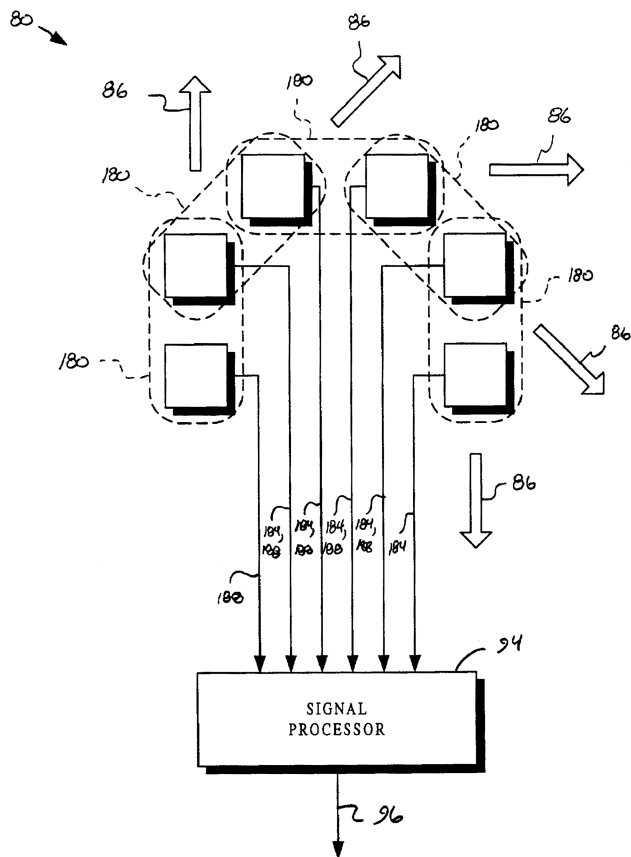
attendant level gating, spectral equalization, and other techniques, can accomplish this.—JME

7,130,430

43.38.Hz PHASED ARRAY SOUND SYSTEM

Jeffrey P. Milsap, Cambridge, Wisconsin
31 October 2006 (Class 381/77); filed 18 December 2001

Adjustable delay is included in the audio signal feed to each loudspeaker of a dense ceiling array. Thus, acoustic signals from all speakers can be made to sum coherently at a given location, and two or more signals can be simultaneously focused at different locations. The basic technology is
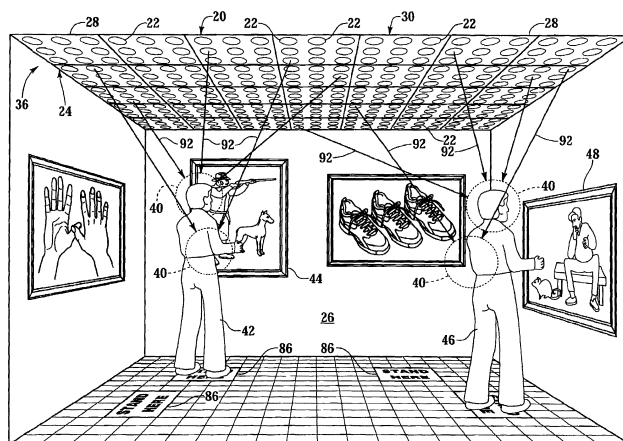
well established, of course, but it seems that no one has previously patented this particular application.—GLA

7,130,441

43.38.Ja SPEAKER SYSTEM HAVING PROCESSING CIRCUITRY

Katsuhiko Tsumori *et al.*, assignors to Fujitsu Ten Limited
31 October 2006 (Class 381/412); filed in Japan 10 November 2000

This patent explains how you can build one loudspeaker for the price of two. A woofer with a heavy cone can generate substantial inertial forces during long excursions. Several commercial loudspeaker designs employ back-to-back woofers to counteract the problem. If the cone of the second woofer is replaced by a moving counterweight **73**, the inertia cancelling
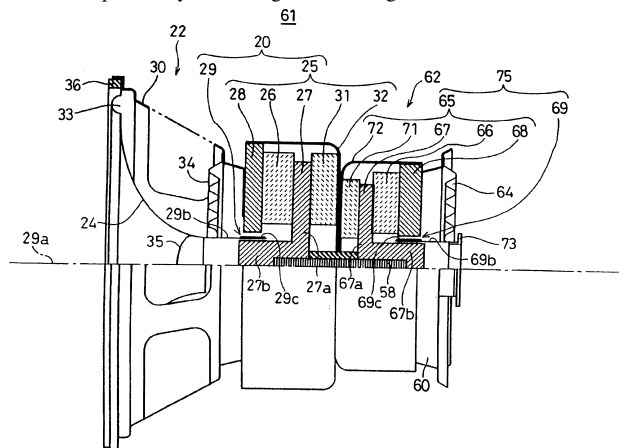
property is retained at the cost of reduced efficiency. In the case of a closed-box system, the efficiency loss might be avoided by driving the second voice coil through a bandpass circuit centered at the system resonant frequency. However, this refinement is not included in the patent's "processing circuitry," which consists of a variable resistor.—GLA

7,133,533

43.38.Ja PASSIVE ACOUSTIC RADIATING

Geoffrey C. Chick *et al.*, assignors to Bose Corporation
7 November 2006 (Class 381/349); filed 21 July 2003

The attached figure shows a section through a device intended to radiate sound while vibrating little. The device consists of two chambers **121A** and **121B**, interconnected rigidly via elements **133** and **135**. Elements **138A** and **138B** are resiliently supported "passive" sound radiators that move in
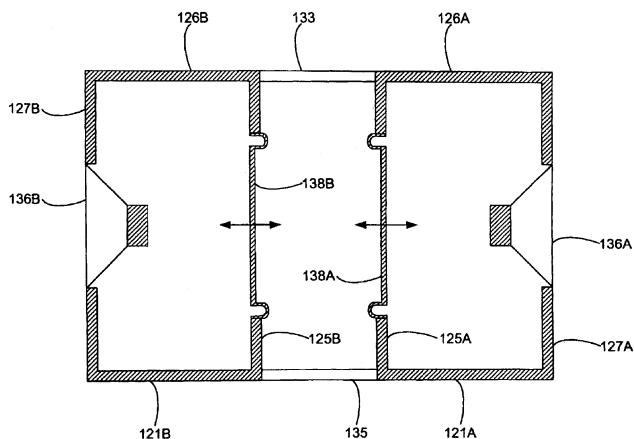
response to pressures in the two chambers. If loudspeakers **136A** and **136B** are driven from the same source, then the passive radiators vibrate in opposite phase so that their inertial forces cancel each other, while they contribute to the sound radiation from the sides of the device (out of the plane of the paper).—EEU

7,136,501

43.38.Ja ACOUSTICALLY ENHANCED ELECTRO-DYNAMIC LOUDSPEAKERS

John F. Steere *et al.*, assignors to Harman International Industries, Incorporated
14 November 2006 (Class 381/399); filed 19 January 2005

This patent describes an improved planar loudspeaker using a "printed circuit" voice coil **1208** bonded directly to the diaphragm. A shorted turn
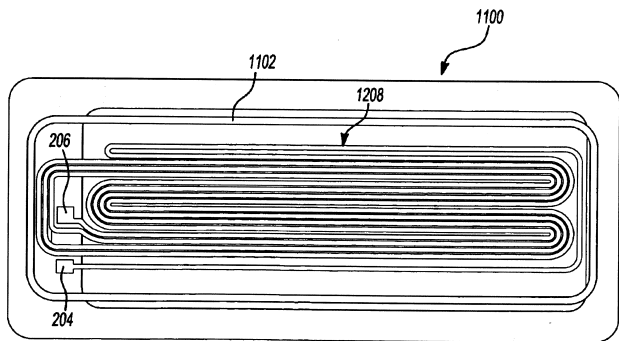
1102 is located along a fringe zone at the edge of the diaphragm to damp edge resonances.—GLA

7,142,685

43.38.Ja ADJUSTABLE LOUDSPEAKER

Dennis Hugh Barnes, assignor to DEI Headquarters, Incorporated
28 November 2006 (Class 381/396); filed 27 August 2003

During design development or for lab demonstrations, it might be convenient to adjust the magnetic flux density in the voice coil gap of a loudspeaker. In the speaker shown, flux density can be adjusted mechani-

cally through a system of slots and shims that introduce additional losses in the magnetic circuit. A much easier, time-tested approach is to use a field coil instead of a permanent magnet and vary the current through the coil.—GLA
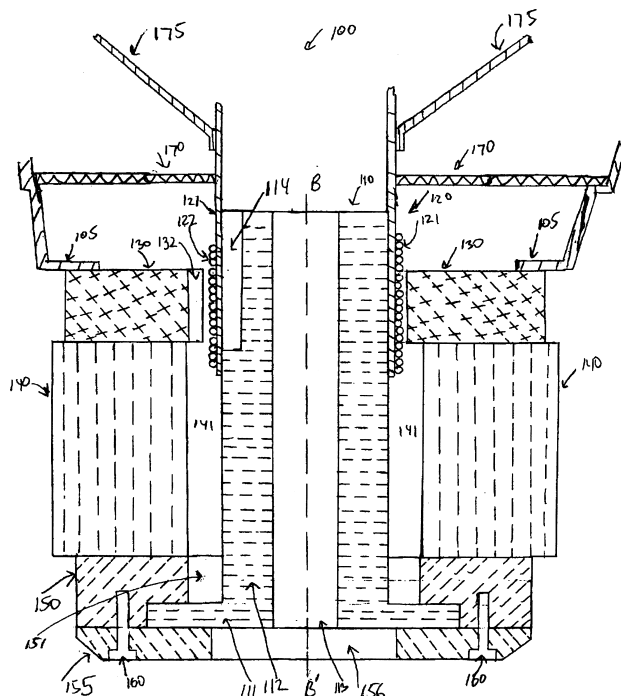
7,142,687

43.38.Ja ELECTROACOUSTIC CONVERTER

Akito Hanada, Yahatanishi-ku, Kitakyashu-shi Fukuoka, Japan
28 November 2006 (Class 381/421); filed in Japan 9 March 2001

This planar-diaphragm loudspeaker has a simple spiral coil bonded to the diaphragm. As in other such designs, the coil is driven by a squished-out magnetic field provided by a partially open magnetic assembly, allowing
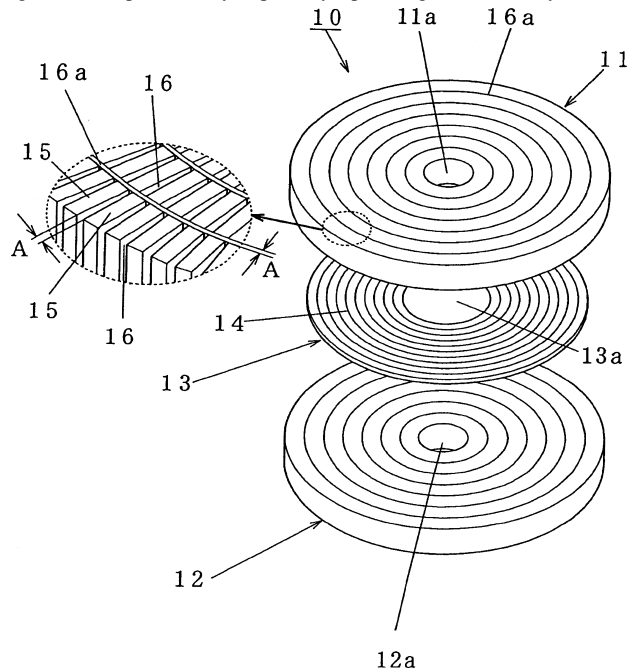
sound to pass through the openings. In this case, front and rear assemblies are used and the direction of magnetization of individual rings varies in relation to their diameters.—GLA

7,113,607

43.38.Md LOW FREQUENCY FEEDBACK CONTROLLED AUDIO SYSTEM

Joe H. Mullins, Albuquerque, New Mexico
26 September 2006 (Class 381/96); filed 2 September 1999

This patent is a brief synopsis, with the required patentese, of the peer reviewed article ''A low-frequency feedback-controlled audio system'' by
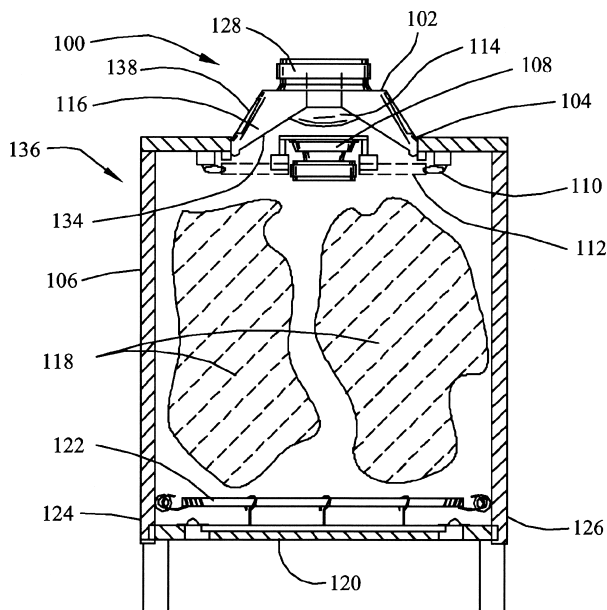
the inventor, published in J. Acoust. Soc. Am. **106**(2), 739 (1999). The inventor references important work by D. B. Keele and M. R. Gander in both the article and the patent.—NAS

7,129,828

43.38.Rh ENCAPSULATED SURFACE ACOUSTIC WAVE SENSOR

James D. Cook, assignor to Honeywell International Incorporated
31 October 2006 (Class 340/447); filed 20 July 2004

All-quartz wireless tire air-pressure and temperature sensor **100** is interrogated via rf antenna **106**–**110**. Tire pressure **135** is transmitted to SAW
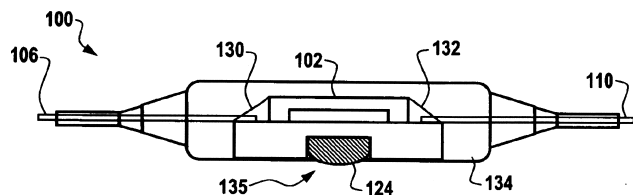
102 substrate **134** which is also active as a diaphragm under incompressible gel **124**. SAW interior volume **130**–**132** is filled with gas at a reference pressure.—AJC

7,132,778

43.38.Rh SURFACE ACOUSTIC WAVE MODULATOR

Paul A. Nysen and Halvor Skeie, assignors to X-Cyte, Incorporated
7 November 2006 (Class 310/313 D); filed 20 August 2003

The patent authors claim a SAW rf tag system comprising a folded SAW path along which many information modulators are located. Rf interrogation energy picked up by an antenna (not shown) travels down bus bar **702**, which inductively couples to SAW transducers **703**–**709**. SAW energy
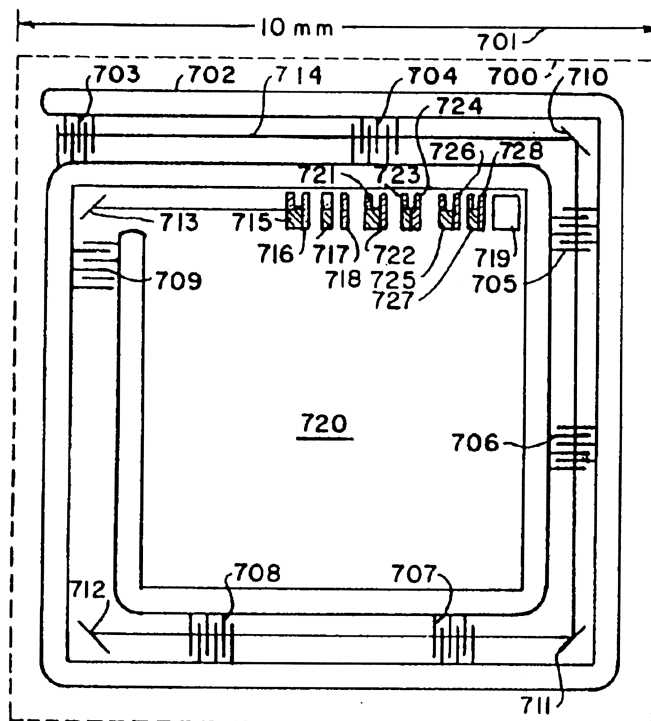
will travel on path **714**, being folded by corner reflectors, to reach folding end reflectors **713**–**728**, which may be broadband, or may be frequency selective. Tag information is encoded as the status of each reflector. Reflected energy is returned along the same acoustic and rf paths. The claims include quadrature amplitude modulation.—AJC

7,132,779

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Hajime Kando, assignor to Murata Manufacturing Company, Limited
7 November 2006 (Class 310/313 R); filed in Japan 24 July 2001

To achieve higher-frequency operation, the sound propagation velocity in quartz SAW substrate **2** is increased by encapsulating interdigital electrode **4a** in a zinc oxide dielectric film **3a**. To optimize the Q and the reflection coefficient where needed, angle θ should be greater than 30° and

2482   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Reviews of Acoustical Patents   2482
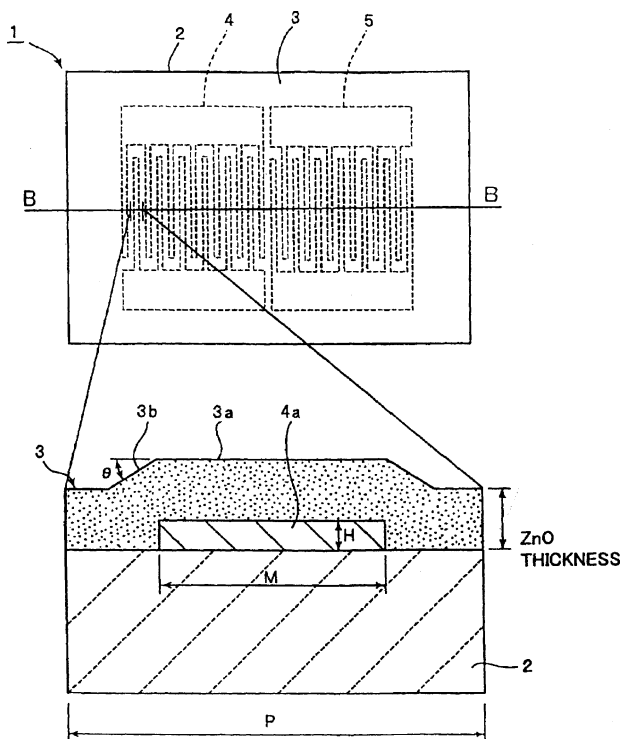
preferably 40° or more. The thickness **H** of aluminum electrode **4a** is made greater than 0.01 wavelengths and the dielectric film density is greater than that of at least one electrode.—AJC

7,135,805

43.38.Rh SURFACE ACOUSTIC WAVE TRANSDUCER

Kazuhiko Yamanouchi, assignor to Nihon Dempa Kogyo Company, Limited

14 November 2006 (Class 310/313 D); filed in Japan 8 April 2003

The functioning paradigm is not clear. This is my interpretation of what I saw and read: It seems to be a 2-pole device, which implies that it could serve as a bandpass filter, a shunting attenuator, or a series impedance. Bus bars **2** and **3** present interspersed electrode fingers **5** and **6** in conjunction with alternate material **4**, such as a dielectric film or other metal. These together serve to impress a voltage into the piezoelectric substrate (under all
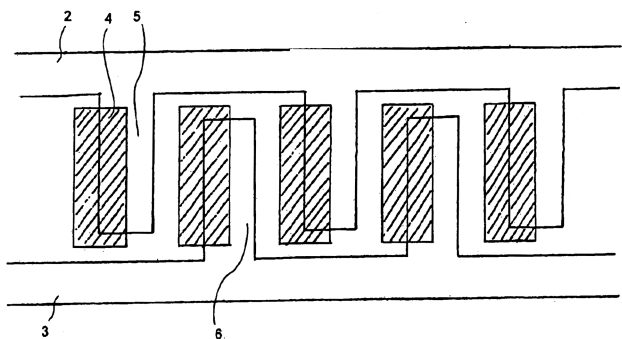
parts) to launch surface acoustic waves. Depending on dimensions and frequency, said surface waves arrive at the opposing electrode either in phase or out of phase; the piezoelectric voltages thus produced either effect constructive (passband) or destructive (stop band) interference. The author claims superior efficiency (no power loss) and simplicity of manufacturing for this method.—AJC

7,141,909

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Takuo Hada et al., assignors to Murata Manufacturing Company, Limited

28 November 2006 (Class 310/313 R); filed in Japan 17 June 2003

To increase resistance to solvents and corrosion while increasing power handling ability, SAW device **1** on piezoelectric substrate **2** has

copper interdigital electrodes **3a**, that are also plated with adhesion-improving metal **3b**, such as a nickel-chromium alloy, and a protective metal layer **3c**, such as aluminum, titanium, or nickel, plus a silicon dioxide overcoat **4**.—AJC

7,130,418

43.38.Si HANDSET FOR COMMUNICATION TERMINAL

Chul Jin Kim et al., assignors to LG Electronics Incorporated

31 October 2006 (Class 379/419); filed in Republic of Korea 24 December 2002

To minimize the pickup of background noise, a cardioid microphone capsule **50** is enclosed in this telephone handset. Sound openings at front

2483   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Reviews of Acoustical Patents   2483
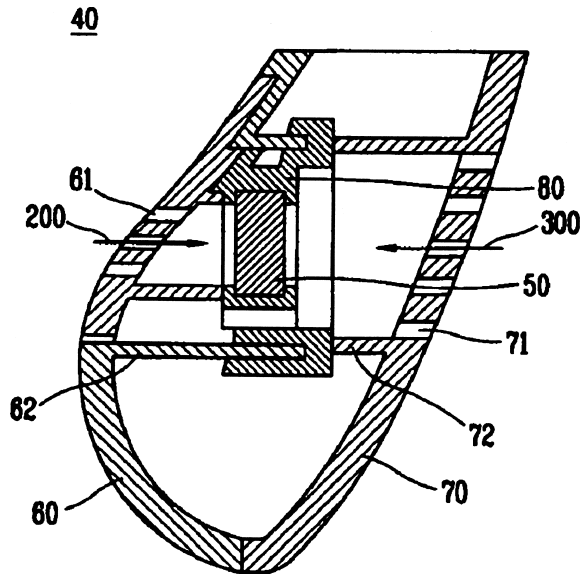
300 and rear **200** maintain the cardioid pattern. Presumably, the microphone's response is tailored to allow for the proximity effect.—GLA

7,136,678

43.38.Si VIBRATING PORTABLE ELECTRONIC DEVICE, METHOD OF VIBRATING A PORTABLE ELECTRONIC DEVICE AND METHOD OF MESSAGING BY VIBRATING A PORTABLE ELECTRONIC DEVICE

Niko Eiden, assignor to Nokia Corporation
14 November 2006 (Class 455/567); filed 26 June 2003

Dual, computer-controlled motors are used to drive offset masses in this cellular phone vibrator. The vibration intensity and frequency can be
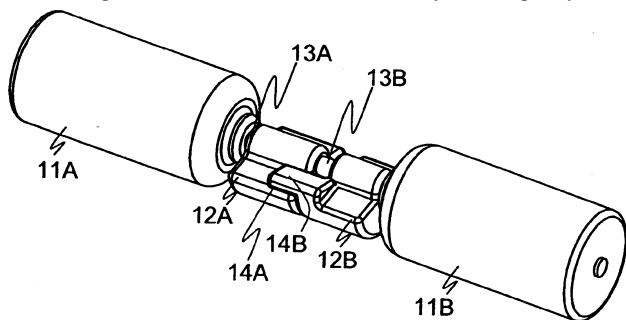
controlled independently, allowing "vibration messages" to be received.—GLA

7,139,394

43.38.Si LEAK-TOLERANT HANDSFREE TELEPHONE

Stéphane Dufosse and Nicolas Janaszek, assignors to TCL Communication Technology Holdings Limited
21 November 2006 (Class 379/428.01); filed in France 20 March 2001

This telephone receiver functions as a miniature bass-reflex loudspeaker. When used as an earphone, the configuration is said to be leak

tolerant. When used as a speakerphone, low frequencies are enhanced and voice quality is improved.—GLA
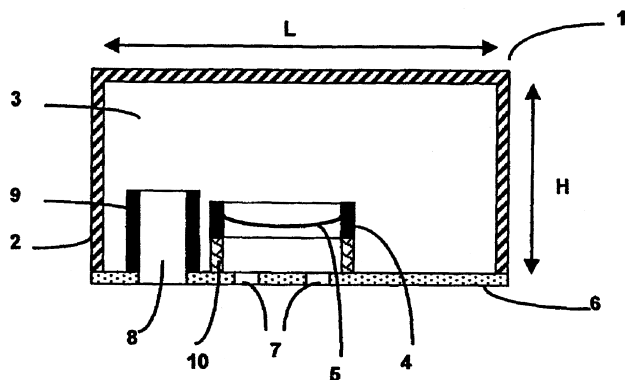
7,136,625

43.38.Si METHOD FOR IMPROVING ACOUSTIC PROPERTIES OF A TERMINAL DEVICE AND A TERMINAL DEVICE

Juha Ylitalo and Antti-Olli Kauppinen, assignors to Flextronics ODM Luxembourg S.A., Luxembourg Zweigniederlassung Schweiz
14 November 2006 (Class 455/90.3); filed in Finland 6 June 2001

The user of a cellular telephone normally holds the device near, but not sealed against, the ear. To cope with variable ear-speaker spacing, some designs provide additional sound channels to maintain a relatively constant
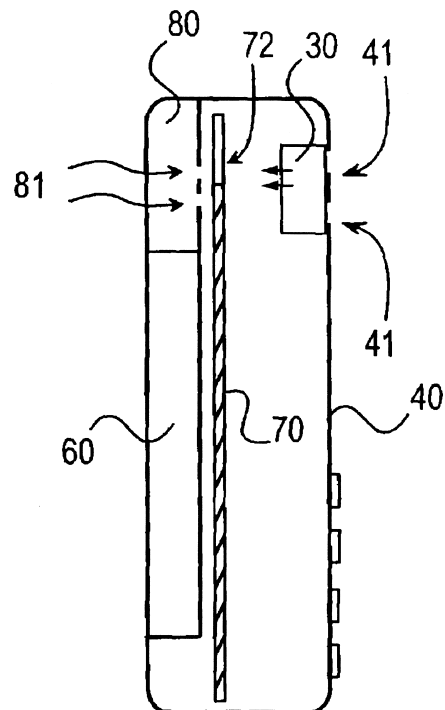
near-field sound level. In this design, rear radiation from the speaker travels through perforations in the antenna board and out through additional perforations in the rear of the case.—GLA

7,142,894

43.38.Si MOBILE PHONE FOR VOICE ADAPTATION IN SOCIALLY SENSITIVE ENVIRONMENT

Fumiko Ichikawa *et al.*, assignors to Nokia Corporation
28 November 2006 (Class 455/569.1); filed 30 May 2003

Users of cellular telephones may unintentionally talk louder than necessary and it is common for voice levels to increase during the course of a conversation. This patent describes circuitry that enables the phone to monitor the background noise level and determine if the user is talking too loud. If so, the side-tone level is increased, and a warning may also be initiated. Apparently, in a small country like Finland, people can afford to be polite.—GLA

7,140,254

43.40.Le APPARATUS FOR GENERATING AND MEASURING BENDING VIBRATION IN A NON-FERROMAGNETIC PIPE WITHOUT PHYSICAL CONTACT

Yoon Young Kim *et al.*, **assignors to Seoul National University Foundation**
28 November 2006 (Class 73/643); filed in Republic of Korea 7 April 2004

For nondestructive testing of a buried or concealed non-ferromagnetic pipe **2**, which may have an outer coating **6**, an electrodynamic bending force is generated via alternating current through coil **3**. The bending moment results from the electrodynamic moment in pipe metal **2** in the presence of
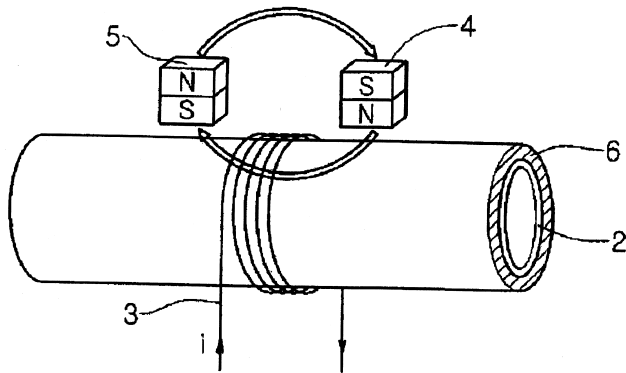
a magnetic field (arrows) from magnets **4** and **5**. The frequency of the alternating current is adjusted to match a bending mode of a pipe **2** of good quality. The damping rate of such bending vibration after current shut-off is observed to ascertain the condition of pipe **2**.—AJC

7,134,710

43.40.Rj AUTOMOTIVE FLOOR PANEL STRUCTURE

Takanobu Kamura and Masaru Chikita, assignors to Mazda Motor Corporation
14 November 2006 (Class 296/193.07); filed in Japan 20 August 2003

This patent describes panel modifications, such as crimps and dimpling, so that the antinodes of the panel's 2×1 mode vibrate in opposite phase and displace nearly equal volumes, reducing the radiation efficiency of this mode.—EEU

7,131,287

43.40.Tm AIR CONDITIONING SYSTEM WITH VIBRATION DAMPENING DEVICE

Robert B. Uselton, assignor to Lennox Manufacturing Incorporated
7 November 2006 (Class 62/296); filed 7 September 2004

This patent describes devices for attenuating vibrations transmitted along pipes that carry refrigerant between elements of air conditioning units. The devices in essence consist of masses that are attached to the pipes at

quarter-wavelength points of the vibrations of concern. The masses are generally of short cylindrical shape and have their axes coincide with the pipe axes. Flexible pipe elements may be employed near the masses.—EEU

7,131,640

43.40.Tm ISOLATOR MOUNT FOR SHOCK AND VIBRATION

Gareth J. Knowles and Bruce Bower, both of Williamsport, Pennsylvania
7 November 2006 (Class 267/136); filed 2 December 2004

This patent describes genetically a wide variety of isolators that may be combined in a variety of ways to achieve good isolation of various types of equipment. The isolators, which may be configured so that they can be snapped together, may consist of combinations of durable fiber-reinforced elastomeric materials, viscoelastic materials, magneto-mechanical and super-elastic alloys, and shape-memory elements. They may also include active control elements consisting of sensing and actuation means, together with the corresponding control circuits. Control algorithms are not discussed.—EEU

7,140,249

43.40.Tm ANGULAR VELOCITY SENSOR

Toru Matsunaga *et al.*, **assignors to Sony Corporation**
28 November 2006 (Class 73/504.12); filed in Japan 30 January 2004

In a typical gyro-type angular velocity sensor, a resiliently supported, rod-shaped element is vibrated at a predetermined resonant frequency, and the Coriolis force resulting from an applied angular velocity is measured. This patent describes the details of such a sensor that is said to be stable, of high quality, and of high reliability.—EEU

7,128,311

43.40.Vn ACTIVE VIBRATION DAMPING ACTUATOR AND ACTIVE DAMPING APPARATUS USING THE SAME

Katsuhiro Goto *et al.*, **assignors to Tokai Rubber Industries, Limited**
31 October 2006 (Class 267/140.14); filed in Japan 26 March 2003

An active damping engine mount **10** is claimed that is easier to manufacture and which will serve as an actuator-damper in conjunction with active vibration control systems. The engine is bolted to **10** via lug **58**. Support is by rubber product **16**. Damping fluid **106** is retained by rubber boot **30**. An orifice plate **96** in conjunction with modulator **86** dynamically
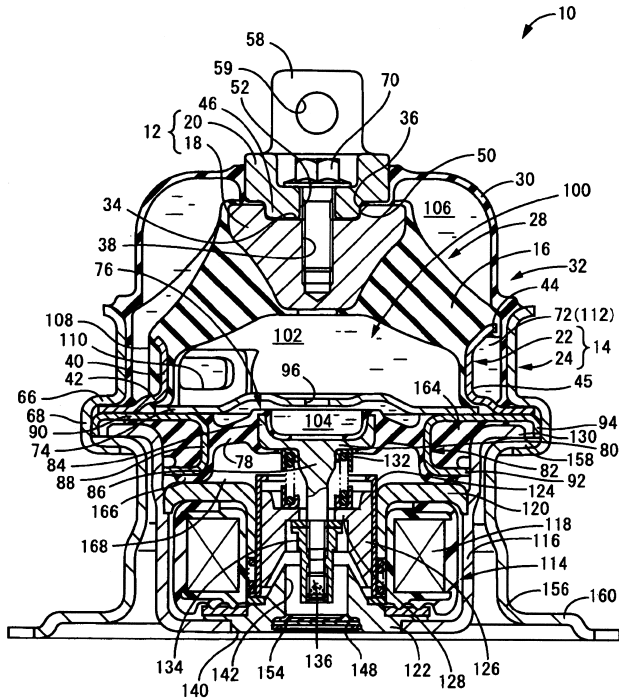
controls fluid flow between chambers **104** and **102**. Fluid flow through and past orifice **96** is modulated by magnetic device **86–134**, driven by electric coil **118**. The author claims improved manufacture, caulking, and sealing means to improve protection against leakage that may arise from installation and adjustment actions.—AJC

7,138,747

43.40.Vn DAMPING AND STABILIZATION FOR LINEAR MOTOR STAGE

James F. Smith *et al.*, **assignors to Anorad Corporation**
21 November 2006 (Class 310/317); filed 29 July 2004

Gantry beams that carry tools or optical equipment tend to vibrate when they are being positioned. In order to reduce their lateral vibrations, they are provided with sensors for measuring the vibrations in selected areas and with actuators that are driven in response to signals from a controller.—EEU

7,133,801

43.40.Yq SYSTEM AND METHODOLOGY FOR VIBRATION ANALYSIS AND CONDITION MONITORING

Limin Song, assignor to Exxon Mobil Research and Engineering Company
7 November 2006 (Class 702/145); filed 12 January 2005

This patent addresses a method that employs adaptive signal processing for the precise determination of the rotational speed of a rotating machine from time-based vibration data. Digitized sample sequences of vibra-

---

tion data are acquired directly from a transducer mounted on the machine, doing away with the need for separate speed measurement systems.—EEU

7,131,315

43.40.Yq METHODS AND APPARATUS FOR REDUCING VIBRATION RECTIFICATION ERRORS IN CLOSED-LOOP ACCELEROMETERS

Donny F. Rojo *et al.*, **assignors to Honeywell International Incorporated**
7 November 2006 (Class 73/1.38); filed 27 April 2005

In a closed-loop accelerometer the displacement of a proof mass relative to its housing is measured by suitable sensors and minimized by actuators that are activated via a control circuit; the actuation signals serve as a measure of the acceleration. This patent describes a means for minimizing the vibration rectification error (VRE). It involves applying a known vibration signal to the accelerometer and adjusting a feedback gain until the VRE is below a first threshold value, then applying a random vibration and adjusting a servo system proportional gain until the VRE is below a second threshold value.—EEU

7,134,336

43.40.Yq VIBRATION TYPE ANGULAR VELOCITY SENSOR

Shunji Mase and Kenji Hirano, assignors to Denso Corporation
14 November 2006 (Class 73/504.12); filed in Japan 19 March 2004

This sensor arrangement, intended for use in vehicle skid control or navigation systems, employs two essentially identical miniature devices. Each of these consists of a vibrating element that is subject to Coriolis forces proportional to the sensor's angular velocity, thus providing a measure of the vehicle's angular velocity. The vibrating elements of the two devices are driven in opposite phase and their outputs are combined so that anomalies in the sensing system can readily be detected.—EEU

7,128,281

43.50.Gf MODULAR FUEL INJECTOR WITH A DAMPER MEMBER AND METHOD OF REDUCING NOISE

Yong D. Cho and Michael P. Dallmeyer, assignors to Siemens VDO Automotive Corporation
31 October 2006 (Class 239/585.1); filed 3 June 2004

Fuel injector noise reduction is claimed where impact noise from
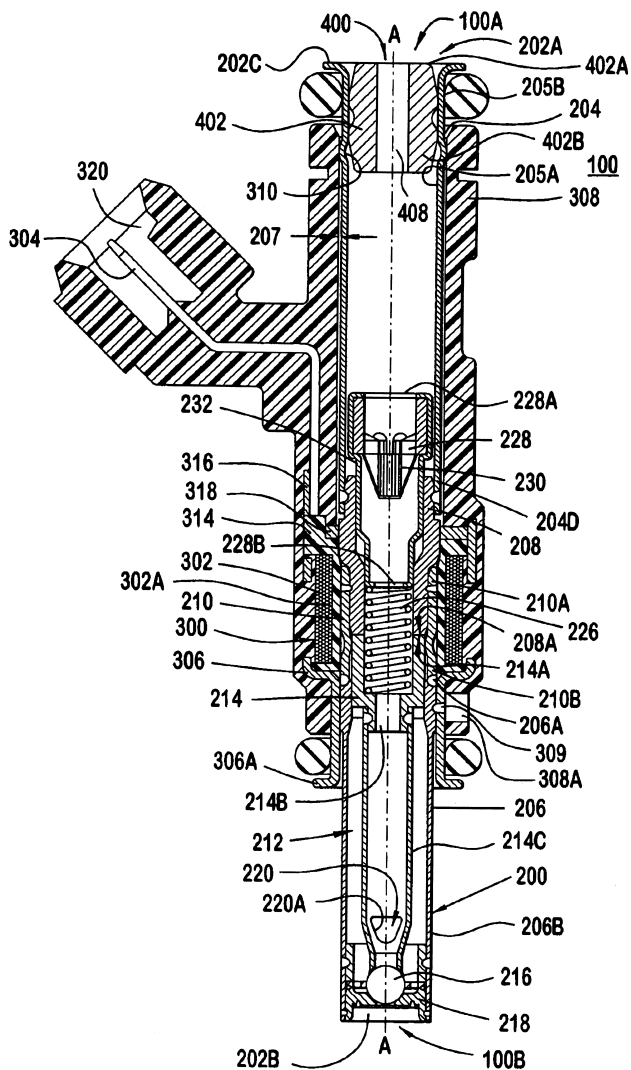
armature **214** is absorbed by damper **402**. A noise reduction of 1 dBA is indicated.—AJC

7,137,477

43.55.Ev SOUND ABSORBERS

Hans-Peter Keller *et al.*, **assignors to Clion Ireland Holding Limited**
21 November 2006 (Class 181/286); filed in Germany 28 May 2003

This sound absorber consists of two interconnected, nonwoven fabrics bonded together with thermoplastic or thermosetting materials, to be used for an engine liner or bulkhead insulation. The patent states that the ratio of the layer thicknesses between the two fabrics should be in the range 3:1–4:l, and that this substantially affects the sound absorption properties. Also, the ratio of densities between the two fabrics should be around 4:1–5:1.—CJR

7,127,858

43.55.Ti INTERIOR WALL AND PARTITION CONSTRUCTION

Derek J. Layfield, assignor to Strawmen, L.P.
31 October 2006 (Class 52/407.1); filed 4 October 2005

This self-supporting wall system eliminates the need for conventional stud framing by using a compressed straw panel as its core. The core has
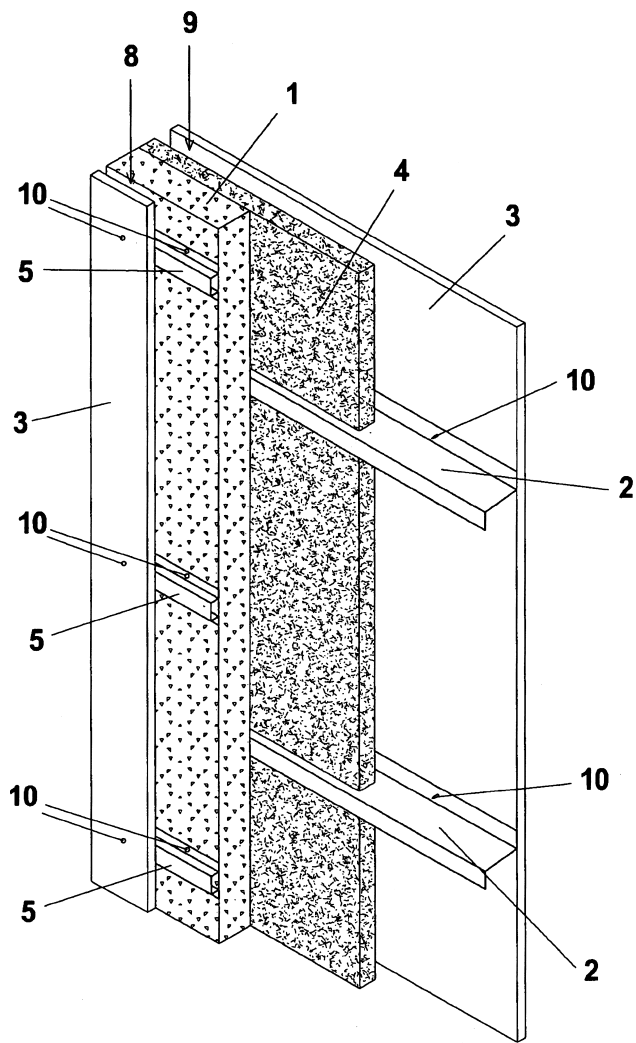
gypsum board attached on each side, with or without resilient channels and insulation in the air cavity between the panel and gypsum board.—CJR

7,139,222

43.60.Dh SYSTEM AND METHOD FOR PROTECTING THE LOCATION OF AN ACOUSTIC EVENT DETECTOR

Kevin Baxter, North Hollywood, California *et al.*
21 November 2006 (Class 367/127); filed 20 January 2005

The patent describes a gun-shot detection system utilizing multiple randomly scattered sensors (microphones) carried by soldiers or vehicles. Each sensor contains a GPS receiver providing location coordinates of that sensor. The claimed innovation is the inclusion of encryption modules for secure transmission of GPS data along with the sensor data, so the position of the sensors, i.e., soldiers or vehicles, cannot be retrieved by an enemy.—DMD

7,130,428

43.60.Mn PICKED-UP-SOUND RECORDING METHOD AND APPARATUS

Toru Hirai *et al.*, **assignors to Yamaha Corporation**
31 October 2006 (Class 381/66); filed in Japan 22 December 2000

Those readers who have been involved with 2-way conferencing systems know the problems of maintaining basic stability in the various noise-

cancelling, source-location-sensitive factors which are entailed here. This patent proposes shifting everything into technical "high gear" by merging both listening and talking environments into a "single acoustic environment," duplicated at both ends. One asks if this can even be done—or if it

even needs to be done. The attached figure gives some idea of the complexities involved. The assignee is the Yamaha Corporation, who has been very busy during the last decade developing convolution-type reverberation generators as well as performance space electronic ambience systems. So far, they have earned high marks in these areas, and I would expect the present effort to be significant.—JME

7,130,437

43.66.Ts COMPRESSIBLE HEARING AID

Paul R. Stonikas *et al.*, assignors to Beltone Electronics Corporation
31 October 2006 (Class 381/322); filed 26 June 2001

A deformable elastomeric hearing aid housing or skin follows the dynamic changes in the shape of a hearing aid wearer's ear canal during talking, chewing, smiling, or swallowing. When deformed, the housing does not attempt to restore itself to the original shape. The skin contains at least
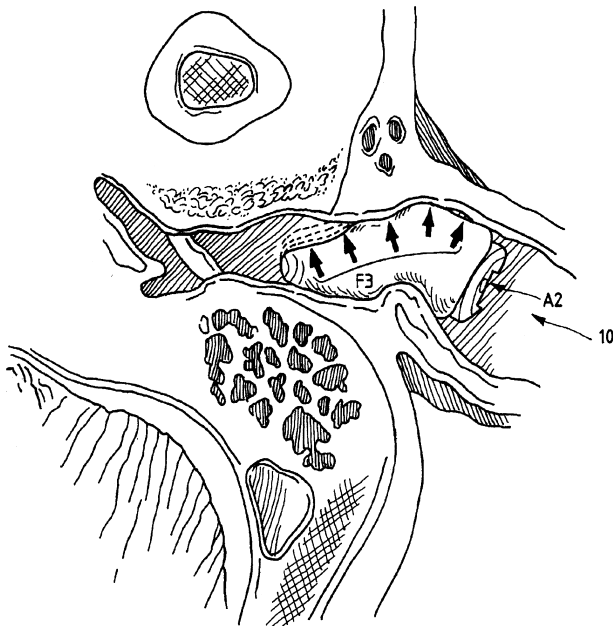
one spine or rib extending axially along an interior surface to provide rigidity when the hearing aid is inserted into the wearer's ear canal. Inside the hearing aid housing, an open cell foam, closed cell foam, or fabric may be incorporated for mechanical isolation of components.—DAP

7,130,694

43.66.Ts PULSE SKIPPING STRATEGY

Andrew W. Voelkel, assignor to Advanced Bionics Corporation
31 October 2006 (Class 607/55); filed 13 December 2002

To improve the representation of weak speech components in a very noisy steady-state background, a probabilistic selection method is used to randomly pick N out of M channels for neural stimulation. A probability for

each of the M channels is computed based on the strength of each channel. Capability is also provided to emphasize specified channels that, for example, carry high-frequency content during transient signals.—DAP

7,136,007

43.66.Ts HEARING-AID REMOTE CONTROL

Josef Wagner, assignor to Phonak AG
14 November 2006 (Class 341/176); filed 18 February 2003

To promote greater flexibility, a remote control for wireless control of a hearing aid is designed in an attachable housing to clip onto or slip onto,
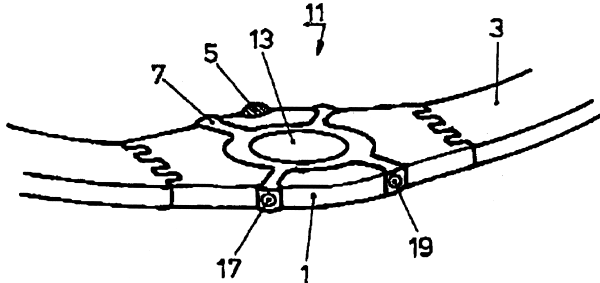
for example, a wristwatch. The remote control can be operated via keys, buttons, or conductive polymers provided in the clips that clamp around the rim of the wrist watch.—DAP

7,139,403

43.66.Ts HEARING AID WITH DIGITAL COMPRESSION RECAPTURE

Garry Richardson and Jerry Wahl, assignors to AMI Semiconductor, Incorporated
21 November 2006 (Class 381/107); filed 8 January 2002

To avoid overdriving the analog-to-digital converter, the input signal to a hearing aid is compressed, which, unfortunately, also produces temporal distortion. A feedforward compression recapture system uses a digital con-

trol signal, which may represent the inverse of the amount of compression applied to the analog input signal multiplied by the compressed digital input signal to digitally restore the original input signal.—DAP

7,139,404

43.66.Ts BTE/CIC AUDITORY DEVICE AND MODULAR CONNECTOR SYSTEM THEREFOR

Jim Feeley and Mike Feeley, assignors to Hear-Wear Technologies, LLC
21 November 2006 (Class 381/330); filed 10 August 2001

The system includes a behind-the-ear (BTE) portion, which contains processing circuitry and a microphone, and a completely in-the-canal (CIC) portion, which contains a speaker and deep-fitting custom ear mold. A connector, having at least one end detachable from the BTE and CIC
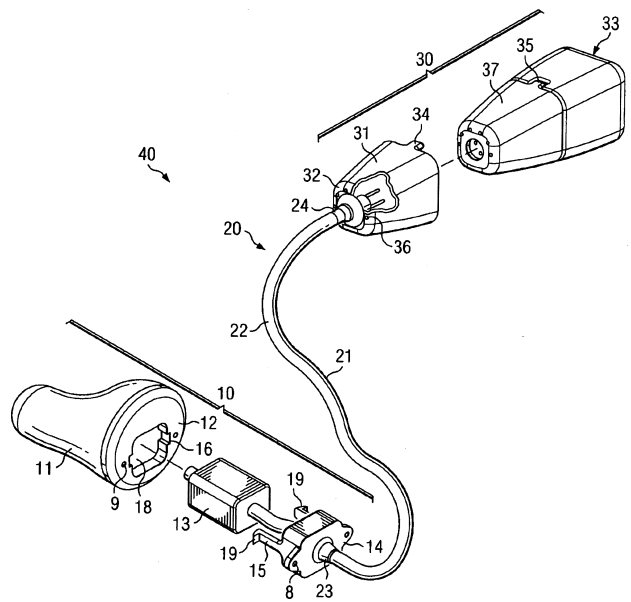
components, electrically couples the BTE and CIC portions. A portion of the connector resides in the ear canal of the wearer and may be used to insert and remove the CIC portion. The speaker may also be detachably coupled to the ear mold.—DAP

7,142,682

43.66.Ts SILICON-BASED TRANSDUCER FOR USE IN HEARING INSTRUMENTS AND LISTENING DEVICES

Matthias Müllenborn and Aart Z. van Halteren, assignors to Sonion Mems A/S
28 November 2006 (Class 381/322); filed 20 December 2002

A silicon-based transducer assembly is packaged on or within a manually movable structure associated with a hearing aid. For example, a silicon-based MEMS microphone array and associated A/D converter, amplifier, digital signal processor—all on a single chip—may be located just under-

neath the outside surface of a removably hinged hearing aid battery drawer. A hole in the battery drawer gives the microphone array access to the environment. There are many other claims, including that the hearing aid may be disposable and may contain a wireless interface for programming or for communication with another hearing aid.—DAP
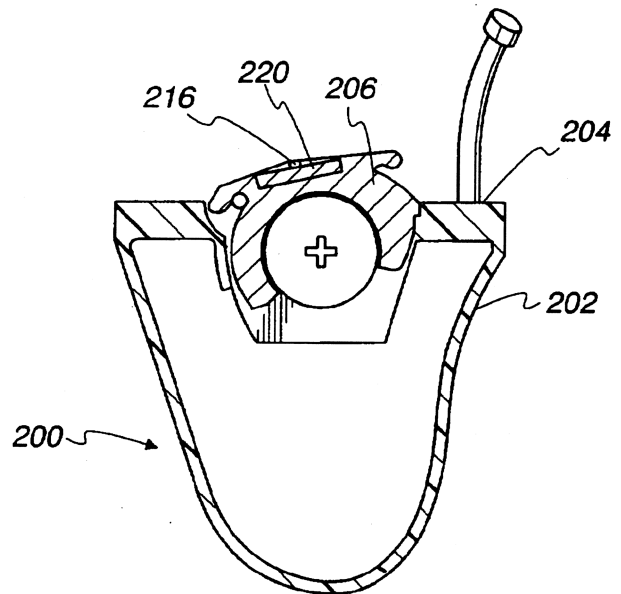
7,141,014

43.66.Ts CUSHIONING DEVICE FOR USE WITH A HEARING AID

Oleg Saltykov, assignor to Siemens Hearing Instruments, Incorporated
28 November 2006 (Class 600/25); filed 30 May 2002

A pliant substance within a ring volume that partially encircles the distal tip of a custom hearing aid is displaced into and out of an attached reservoir. An air volume within the pliant substance is compressed and de-compressed by the battery door closing and opening, respectively, to adjust
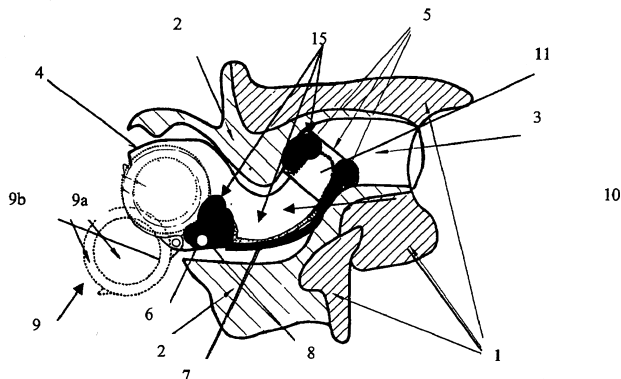
the shape of the ring volume to the contour of the inner portion of the wearer's ear canal. The result is a flexible cushioning device that positions the hearing aid in the ear canal of the wearer so as to create an acoustical seal.—DAP

7,132,949

43.66.Yw PATIENT MANAGEMENT IN AUTOMATED DIAGNOSTIC HEARING TEST

Jeffrey S. Harrison *et al.*, assignors to Tympany, Incorporated
7 November 2006 (Class 340/573.1); filed 16 May 2003

A method is proposed for computer-based, self-administered, multilin-gual hearing tests which can be performed without a sound isolation cham-ber. Several alerting mechanisms are included, for example, wireless paging
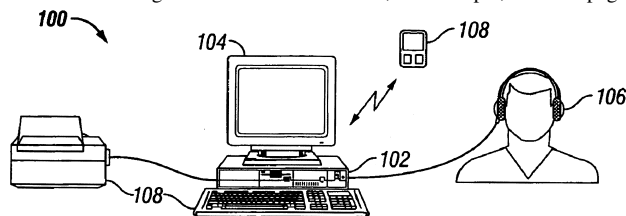
or text messaging, in case operator assistance is needed by the patient. An ambient noise threshold-shift detector continuously monitors whether the ambient noise level exceeds the level specified by an ANSI standard and measures noise duration.—DAP

7,133,715

43.66.Yw HEARING EVALUATION DEVICE WITH NOISE DETECTION AND EVALUATION CAPABILITY

Matthijs P. Smits and Bryan P. Flaherty, assignors to Natus Medical Incorporated
7 November 2006 (Class 600/544); filed 7 January 2000

In extracting the auditory brainstem response (ABR) wave form, the evaluation is paused if the nonphysiological noise contribution is deemed to exceed a preset threshold. To determine if an ABR is present, the collected

wave forms are digitized and compared to a reference ABR wave form by calculating the sum of the polarities at each measured point in time. If the observed polarity-sum distribution is statistically different than that expected from random noise, the ABR is deemed present.—DAP

7,141,022

43.66.Yw METHOD FOR ALIGNING DERIVED-BAND ABR RESPONSES BASED ON INTEGRATION OF DETRENDED DERIVED-BAND ABRS

Manuel Don and Curtis W. Ponton, assignors to House Ear Institute
28 November 2006 (Class 600/559); filed 5 April 2004

Wave V amplitudes of repeated measures of the auditory brainstem response (ABR) are traditionally time aligned via their peaks, summed, and compared to a threshold value to predict whether small intracanalicular tu-mors are present. Suggested here is a method of alignment that eliminates
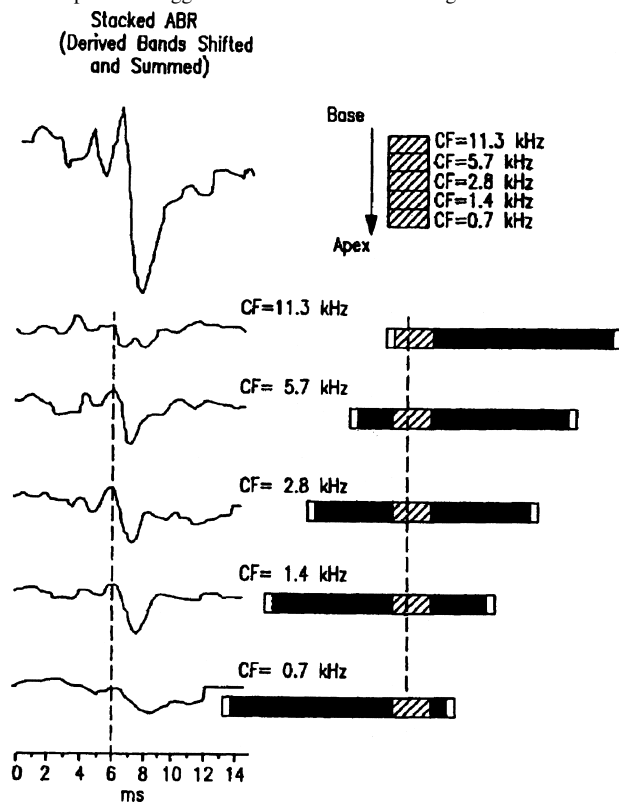
subjective judgment using zero crossings of Wave V prior to summing the shifted ABRs. Each of the derived-band ABRs is fit to a linear equation, which, after subtracting out dc shifts, is integrated to identify a peak and peak displacement for each ABR. The normalized ABRs are then shifted by the identified displacement values.—DAP

7,011,525

43.71.Ft ENCODING SYSTEM COMBINING LANGUAGE ELEMENTS FOR RAPID ADVANCEMENT

Roberto G. Mejia, assignor to Literacy S.T.A.R.
14 March 2006 (Class 434/167); filed 9 July 2003

Fifty years ago, children learned to build houses and towers and such of little wooden blocks with letters on the sides. This patent makes the case that the child can also learn some phonetics while learning the written

2490   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Reviews of Acoustical Patents   2490

(S a t ur d ay)

letters. Each side of a block would have a picture which supposedly invokes some specific word or sound. The letters used to spell that sound are then also displayed around the picture.—DLR

7,120,575

43.72.Ar METHOD AND SYSTEM FOR THE AUTOMATIC SEGMENTATION OF AN AUDIO STREAM INTO SEMANTIC OR SYNTACTIC UNITS

Martin Haase *et al.*, assignors to International Business Machines Corporation
10 October 2006 (Class 704/207); filed in Germany 8 April 2000

A simplistic scheme is outlined for using properties of the fundamental frequency of a speech signal to determine likely prosodic boundaries that would tend to indicate syntactic or semantic boundaries. The properties of the voice pitch to be exploited would include its relative value within the speaker's range and the slope of the pitch track.—SAF

7,124,075

43.72.Ar METHODS AND APPARATUS FOR PITCH DETERMINATION

Dmitry Edward Terez, Millville, New Jersey
17 October 2006 (Class 704/203); filed 7 May 2002

An ingenious approach is suggested here to pitch determination using state-space analysis of the digitized speech signal, viewed as a time series. The signal is converted into a sequence of m-dimensional vectors in a state space, and a frequency histogram is then generated by accumulating total numbers of the closest pairs of vectors with the same time separations into corresponding histogram bins. The peaks in the histogram should then correspond to a fundamental period and its integer multiples.—SAF

7,127,389

43.72.Ar METHOD FOR ENCODING AND DECODING SPECTRAL PHASE DATA FOR SPEECH SIGNALS

Dan Chazan and Zvi Kons, assignors to International Business Machines Corporation
24 October 2006 (Class 704/205); filed 13 September 2002

A method is described for encoding the spectral phase of small speech segments, with the goal of improving the results of concatenative synthesis, since that technique is plagued with phase mismatch across segment transitions. The core method involves computing the "linear phase term" that minimizes "the total variation of the complex spectrum divided by the square root of its absolute value." Successive segments can then be aligned using their complex spectra supplemented with the phase information. An alternative method involving a Hilbert filter is also discussed.—SAF

7,016,505

43.72.Dv ROBOT ACOUSTIC DEVICE

Kazuhiro Nakadai *et al.*, assignors to Japan Science and Technology Agency
21 March 2006 (Class 381/71.1); filed in Japan 30 November 1999

This patent deals with the issue of noise pickup by the microphones which serve as the hearing devices for a humanoid robot. A human-like arrangement is assumed, in which the hearing-sense microphones **16** are
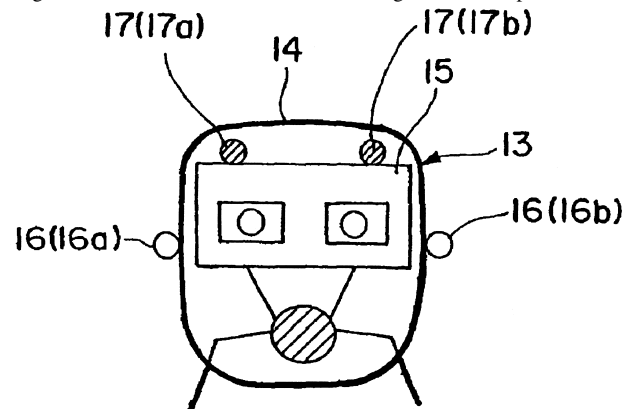
mounted on the lateral surfaces of the head. Additional internal microphones **17** are positioned so as to pick up internal noises, such as motors, etc. These latter microphones then serve as noise-cancelling reference pickups.—DLR

7,117,145

43.72.Dv ADAPTIVE FILTER FOR SPEECH ENHANCEMENT IN A NOISY ENVIRONMENT

Saligrama R. Venkatesh and Alan M. Finn, assignors to Lear Corporation
3 October 2006 (Class 704/200); filed 19 October 2000

A sophisticated approach is described in this patent for adaptively filtering noise from speech in a moving vehicle. The methods are described as implemented in a "cabin communication system." The core of the scheme involves an interesting weighted least squares modification to the basic Weiner-Hopf equation defining a causal filter in the frequency domain. The authors note that a naive approach to adaptively solving the Weiner-Hopf equation yields an adaptive filter that is itself noisy because of sudden spectral changes from frame to frame. By framing the solution as a weighted least squares system, a variable amount of smoothing based on past frequency-bin values can be employed.—SAF

7,117,148

43.72.Dv METHOD OF NOISE REDUCTION USING CORRECTION VECTORS BASED ON DYNAMIC ASPECTS OF SPEECH AND NOISE NORMALIZATION

James G. Droppo *et al.*, assignors to Microsoft Corporation
3 October 2006 (Class 704/228); filed 5 April 2002

A noise-reduction technique is described, wherein a feature vector extracted from a noisy signal is used in a variety of suggested procedures to select a correction-feature vector in a way that is sensitive to the dynamic aspects of the signal. These correction-feature vectors have to be initialized by a training procedure that optimizes the grouping of vectors into mixture components.—SAF

7,013,278

43.72.Ja SYNTHESIS-BASED PRE-SELECTION OF SUITABLE UNITS FOR CONCATENATIVE SPEECH

Alistair D. Conkie, assignor to AT&T Corporation
14 March 2006 (Class 704/260); filed 5 September 2002

Something is missing here. A speech synthesizer would be used to generate a massive database of triphones, which are then stored for later use by a real-time concatenative phoneme synthesizer. But the whole point of concatenative synthesis is to have access to the quality possible with human speech. If the tables are synthesized, what is the point?—DLR

7,016,840

43.72.Ja METHOD AND APPARATUS FOR SYNTHESIZING SPEECH AND METHOD APPARATUS FOR REGISTERING PITCH WAVE FORMS

Ryo Mochizuki *et al.*, assignors to Matsushita Electric Industrial Company, Limited
21 March 2006 (Class 704/258); filed in Japan 18 September 2000

The authors of this patent have noticed that when the speech analysis frame advances by a fixed number of samples, then, during a voiced wave form, each successive glottal pulse will be captured with a different phase. This phase variation makes the task of comparing and combining wave form segments for synthesis much more difficult. Things go much better if the various captured glottal pulse exemplars are transformed to have similar
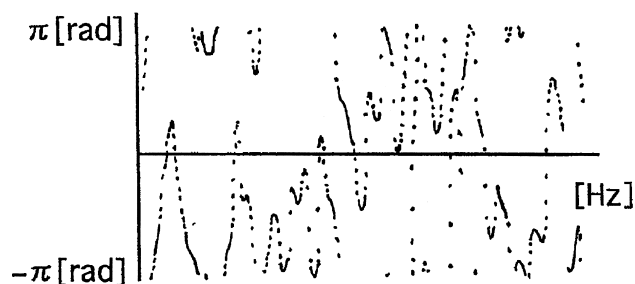
phases. Various ways are presented for doing this phase transformation, the primary method being that the phase components of many Fourier transformed pulses are averaged, the actual phase of a pulse transform then being replaced by the average phase, followed by an inverse transform. In this way, many pulses with different phases are collapsed into a single data record. Exactly how the averaging would be done is not said and there is no mention of modulo 2π issues. Various pitches and various vowel qualities would still be kept in separate data tables.—DLR

7,120,584

43.72.Ja METHOD AND SYSTEM FOR REAL TIME AUDIO SYNTHESIS

Hamid Sheikhzadeh-Nadjar *et al.*, assignors to AMI Semiconductor, Incorporated
10 October 2006 (Class 704/266); filed in Canada 22 October 2001

This patent resurrects a variation on diphone-based text-to-speech synthesis, a technique which has fallen out of favor in recent years due to its notoriously poor quality. Nevertheless, the authors note that the more modern approaches have rather excessive storage and computation requirements that may leave room for tweaking of the simpler diphone methodology to allow real-time synthesis on small computers as described in the patent. Provisions for compression and decompression of the speech units are discussed, as well as concatenation using circular-shift pitch-synchronous overlap add. It is not entirely clear just what aspects of the patent constitute novel inventions.—SAF

7,124,083

43.72.Ja METHOD AND SYSTEM FOR PRESELECTION OF SUITABLE UNITS FOR CONCATENATIVE SPEECH

Alistair D. Conkie, assignor to AT&T Corporation
17 October 2006 (Class 704/266); filed 5 November 2003

This patent notes that current "large database" approaches to concatenative speech synthesis generally use approximately phone-sized units that are treated as if they were possible in all combinations, which is far from ideal. The suggestion offered is to preselect a universe of possible phonemes based upon their triphone context. Accordingly, "a triphone database is created where, for any given triphone context required for synthesis, there is a complete list, precalculated, of all the units (phonemes) in the database that can possibly be used in that triphone context."—SAF

7,127,396

43.72.Ja METHOD AND APPARATUS FOR SPEECH SYNTHESIS WITHOUT PROSODY MODIFICATION

Min Chu and Hu Peng, assignors to Microsoft Corporation
24 October 2006 (Class 704/258); filed 6 January 2005

The extreme of concatenative speech synthesis would be to store speech segments as spoken in the desired prosodic environment, to obviate the need for prosody modification (a notorious cause of poor quality speech). The problem is then, of course, storing all those units and selecting them for the synthesis. This patent describes new techniques for efficiently selecting stored speech segments having desired prosodic properties, exploiting single and multitier Viterbi search.—SAF

7,127,397

43.72.Ja METHOD OF TRAINING A COMPUTER SYSTEM VIA HUMAN VOICE INPUT

Eliot M. Case, assignor to Qwest Communications International Incorporated
24 October 2006 (Class 704/260); filed 31 May 2001

"I have a new word that I do not know how to pronounce. Do you have time to listen to my question?" This excerpt is from a computer-human dialogue envisioned in the patent, as a means by which a computer could exploit a human teacher to train its own pronunciation database. The computer would ask its human to pronounce an unknown word spelled on the screen, and incorporate the information. The whole concept is patented as a

2492   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Reviews of Acoustical Patents   2492

generality, without any specific details of how a speech recognizer could be used to provide speech segments of a sort useful to a speech synthesizer.—SAF

7,139,709

43.72.Ja MIDDLEWARE LAYER BETWEEN SPEECH RELATED APPLICATIONS AND ENGINES

Philipp Heinz Schmid *et al.*, assignors to Microsoft Corporation
21 November 2006 (Class 704/258); filed 29 December 2000

Communication between speech synthesis application programs and speech synthesis engines is facilitated by providing mechanisms in between them to parse input data into text fragments, optimize data formats, configure the speech recognition engine, handle event notification, enable gram-

mar manipulation, manage data output to an audio device, and manage multiapplication, multivoice mixing. Alert priority requests are recognized for a text-to-speech engine to speak a particular message prior to others.—DAP
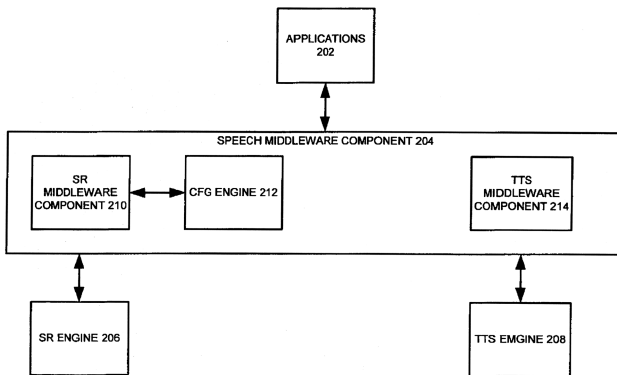
7,013,265

43.72.Ne USE OF A UNIFIED LANGUAGE MODEL

Xuedong D. Huang *et al.*, assignors to Microsoft Corporation
14 March 2006 (Class 704/9); filed 3 December 2004

This patent makes the case that prior speech recognition systems often used various types of grammars, *n*-gram word sets, context-free grammars, etc., compiling *n*-best hypotheses separately, based on each grammar type. The problem is that good guesses from one grammar may be rejected be-

cause they did not make the *n*-best list of some other grammar. That problem would be solved by combining several types of grammar into a single, unified structure.—DLR
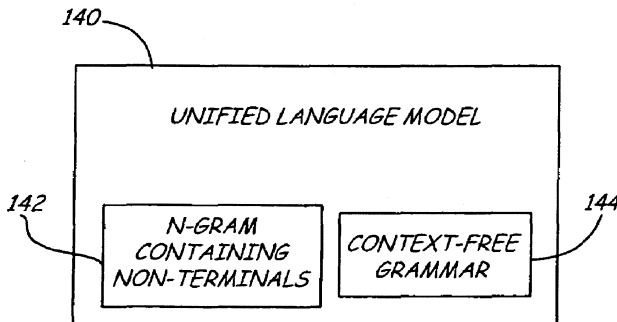
7,013,275

43.72.Ne METHOD AND APPARATUS FOR PROVIDING A DYNAMIC SPEECH-DRIVEN CONTROL AND REMOTE SERVICE ACCESS SYSTEM

James F. Arnold *et al.*, assignors to SRI International
14 March 2006 (Class 704/244); filed 28 December 2001

Here is yet another patent based on the idea that a powerful server can run bigger and more elaborate speech recognition processes than a small mobile device. A recognition system running on the small device should thus be able to make use of services provided by the central machine. In the hopes of carving out new ground in this heavily populated field, the claims list a number of specific services that could thus be provided.—DLR

7,016,830

43.72.Ne USE OF A UNIFIED LANGUAGE MODEL

Xuedong D. Huang *et al.*, assignors to Microsoft Corporation
21 March 2006 (Class 704/9); filed 3 December 2004

This patent is nearly identical with United States Patent 7,013,265, reviewed above, except that the claims have been rewritten.—DLR

7,016,836

43.72.Ne CONTROL USING MULTIPLE SPEECH RECEPTORS IN AN IN-VEHICLE SPEECH RECOGNITION SYSTEM

Shoutarou Yoda, assignor to Pioneer Corporation
21 March 2006 (Class 704/233); filed in Japan 31 August 1999

This speech recognizer would include a special audio-input processing section capable of dealing with input from multiple microphones. This capability is used in two different ways. First, multiple speakers can be easily accomodated in the case where different microphones **M1–M6** are situated
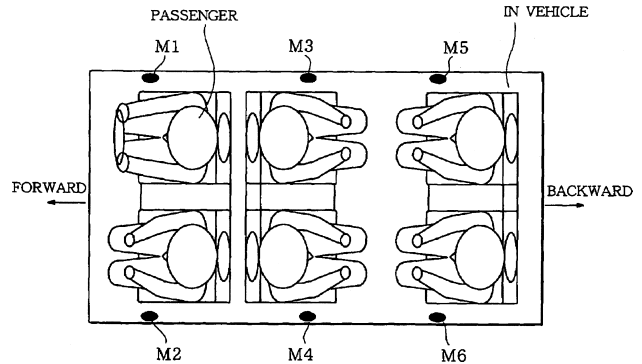
so as to better pick up the various individuals. Second, the input streams from additional microphones can be used during the input from a particular speaker by using the additional signals as noise cancelling inputs. The general idea is what I always imagined the audio system would be like on the bridge of the starship Enterprise.—DLR

7,130,401

43.72.Ne SPEECH TO TEXT CONVERSION SYSTEM

Fred D. Rampey and James M. MacMillan, assignors to Discernix, Incorporated
31 October 2006 (Class 379/142.14); filed 9 March 2004

A system automatically routes speech communication and a caller ID to a communications server which accesses a particular user account. The speech information is then converted to text with a speech recognition sys-

2493   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Reviews of Acoustical Patents   2493

tem configured for the particular user accessed via the caller ID. The resulting text information is transmitted automatically to a destination using routing information stored in the user account.—DAP

7,133,826

43.72.Ne METHOD AND APPARATUS USING SPECTRAL ADDITION FOR SPEAKER RECOGNITION

Xuedong Huang and Michael D. Plumpe, assignors to Microsoft Corporation
7 November 2006 (Class 704/240); filed 24 February 2005

This patent introduces a method to enhance the performance of speaker recognition systems. Normally, state-of-the-art speaker recognition systems employ spectral abstraction techniques to remove noise from the test and training data. According to the patent, spectral subtraction has two drawbacks, a) after subtraction, some frequency components may have negative energy values and b) spectral subtraction doesn't remove all noises, which sometimes causes mismatches between training and test data. To overcome these issues, this invention suggests adding noise rather than subtracting the noise. The method matches the noise in training data to the noise in test data using spectral addition.—AAD

7,136,459

43.72.Ne METHODS AND APPARATUS FOR DATA CACHING TO IMPROVE NAME RECOGNITION IN LARGE NAMESPACES

Robert S. Cooper *et al.*, assignors to Avaya Technology Corporation
14 November 2006 (Class 379/88.03); filed 5 February 2004

A voice dialing system has a directory containing telephone numbers that may be called by a user and a voice recognition module to analyze the voice input from the user, examining the directory for matched entry candi-

dates. The list of candidates is ranked by match confidence and each entry is compared to entries in a called-party cache for each user, considering all such entries that have been assigned a probability of being called by the user.—DAP

7,136,462

43.72.Ne NETWORK SPEECH-TO-TEXT CONVERSION AND STORE

Mariana Benitez Pelaez and Charu Verma, assignors to Lucent Technologies Incorporated
14 November 2006 (Class 379/88.14); filed 15 July 2003

A system converts speech to text if the called user device profile indicates that text is desired. For those users that will send voice, the voice messages may be converted into text.—DAP

7,136,815

43.72.Ne METHOD FOR VOICE RECOGNITION

Harinath Garudadri, assignor to Qualcomm, Incorporated
14 November 2006 (Class 704/255); filed 12 November 2003

A method is suggested for improved storage of voice recognition templates. Voice recognition models are created from the extracted features of

test utterances. The models are compressed for reduced storage and then expanded for comparison against the features of a test utterance. Compression may be performed using A-law coding.—DAP

7,139,707

43.72.Ne METHOD AND SYSTEM FOR REAL-TIME SPEECH RECOGNITION

Hamid Sheikhzadeh-Nadjar *et al.*, assignors to AMI Semiconductors, Incorporated
21 November 2006 (Class 704/243); filed in Canada 22 October 2001

To speed up the speech recognition process, separate processor units are dedicated to inputting speech samples and organizing them into frames, feature extraction via pattern matching, and bin-energy factorization. The processors may operate in parallel and share memory.—DAP

Speech Signal

S2 — Framing — Input-Output Processor (30)

S4 — -Windowing -FFT — WOLA Co-processor (14)

S6 — -Bin Sums -Bin Logs -DCT — DSP Core Processor (20)

S8 — Bin Energy Factorization — WOLA Co-processor (14)

S10 — -Energies -Endpoints — DSP Core Processor (20)

S12 — Training — Off-line

S14 — Recognition — DSP Core Processor (20)
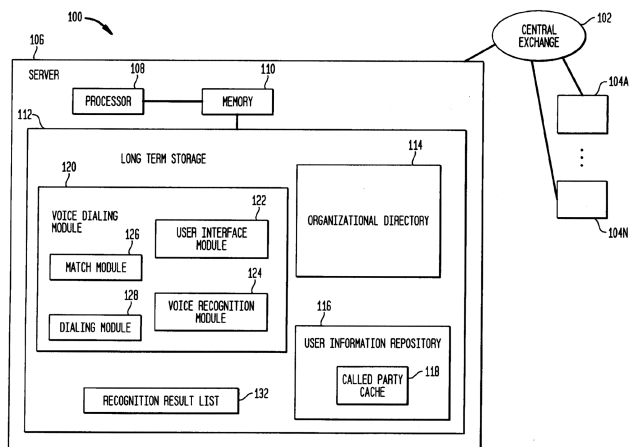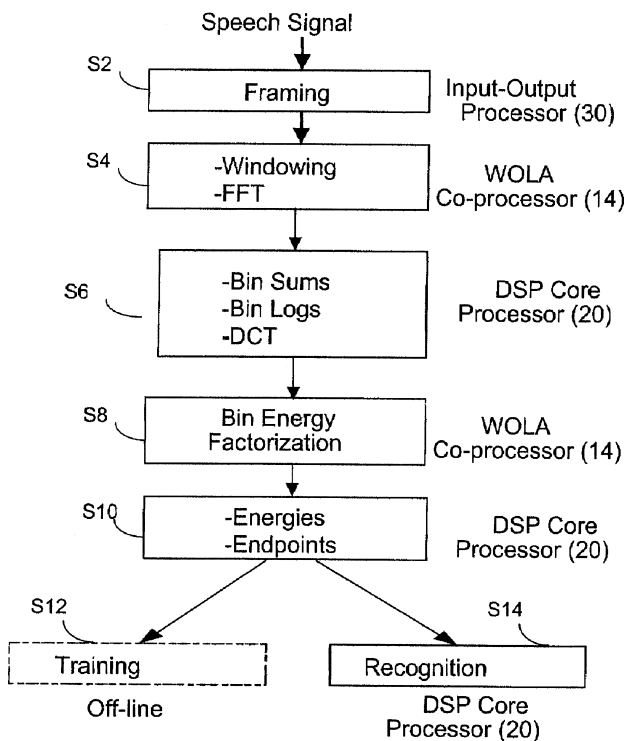
7,139,714

43.72.Ne ADJUSTABLE RESOURCE BASED SPEECH RECOGNITION SYSTEM

Ian M. Bennett *et al.*, assignors to Phoenix Solutions, Incorporated
21 November 2006 (Class 704/270.1); filed 7 January 2005

Speech-based interaction for Internet searches from wireless devices, for example, is enabled by accounting for different processing capabilities of various system components. A distributed-processing speech recognition system utilizes an initial evaluation of the amount of computing resources available at the network server system and specification of an appropriate first set of partial speech recognition operations, while selecting the appropriate engine for the natural language used in a query. After also evaluating the amount of computing resources available at the client device, additional speech recognition operations are performed by the client device to complete the process.—DAP

7,128,712

43.80.Vj ADAPTIVE ULTRASOUND IMAGING SYSTEM

Kenneth Wayne Rigby, assignor to General Electric Company
31 October 2006 (Class 600/447); filed 21 June 2004

The relative motion is estimated between a transducer and a region being imaged. The estimation is used to control a beamformer in which focusing delays are changed according to estimates of sound-speed variation between paths. The variation of path delays can be disabled to avoid image artifacts when the variations in delay are not accurately estimated.—RCW

7,131,948

43.80.Vj METHOD AND APPARATUS FOR INTRAVASCULAR TWO-DIMENSIONAL ULTRASONOGRAPHY

Paul G. Yock, assignor to Scimed Life Systems, Incorporated
7 November 2006 (Class 600/466); filed 22 October 2003

This system uses a catheter that includes a tubular element and an internally housed drive cable that rotates a transducer in the catheter to produce an image in a plane orthogonal to the axis of the catheter.—RCW

7,141,020

43.80.Vj PORTABLE 3D ULTRASOUND SYSTEM

Mckee D. Poland and Martha G. Wilson, assignors to Koninklijke Philips Electronics N.V.
28 November 2006 (Class 600/447); filed 20 February 2002

The ultrasound transducer in this system is two dimensional and includes subarray beamformers. A full array of beamformed signals and envelope detection are performed on a card that plugs into a personal computer (PC). A scan converter produces three-dimensional images that are displayed by the PC.—RCW

7,142,905

43.80.Vj VISUAL IMAGING SYSTEM FOR ULTRASONIC PROBE

Michael H. Slayton and Peter G. Barthe, assignors to Guided Therapy Systems, Incorporated
28 November 2006 (Class 600/427); filed 4 February 2003

A representation is produced of the transducer position during diagnostic or therapeutic treatment. The system also captures acoustic images and other information such as temperature. Information about the transducer position on a patient may be used to image similar planes in the same patient or in subsequent patients. The imaging information can also be used for training.—RCW

LETTERS TO THE EDITOR

Compensating for miter bends in cylindrical tubing (L)

John W. Coltman[a]

3319 Scathelocke Road, Pittsburgh, Pennsylvania 15235

Miter bends in cylindrical tubing perturb wave propagation in a way that may alter the desired resonance properties in musical instruments. The nature of the perturbations and several methods of compensating for them at low frequencies are described. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713669]

I. INTRODUCTION

The air columns of many wind instruments, and some pedal organ pipes, are folded to shorten the instrument. The folds are usually formed by toroidal bends. Nederveen[1] has analyzed the effects of such bends and shown how to compensate them. Toroidal bends are difficult to produce. Amateur wind instrument makers can much more easily make mitered bends, where the tubing is cut at 45 deg and reassembled to make a 90-deg bend. Dequand *et al.*[2] have calculated and measured the reflection coefficients of such bends, but have provided no information on how they may be compensated. The present letter presents methods for compensating miter bends in cylindrical tubing.

II. MITER BEND EFFECTS

A miter bend section is defined by the two planes perpendicular to the axes and intersecting at the inside inner corner of the bend. The physical length of the section is measured along the centerlines and is equal, for a 90-deg bend, to the tubing ID, its inner diameter. The volume, and therefore the compliance, of a mitered section is unchanged from that of a straight section of the same length. Vibratory motion, however, takes a shortcut around the bend, the flow lines being more concentrated toward the inside corner. Thus the inertance is reduced, and the characteristic impedance of the section is altered. When such a bend is located at a pressure maximum in a resonant tube, there is no effect, but if it is located at a pressure minimum the section appears acoustically shortened. The measurements of Dequand *et al.*[2] on a 90-deg bend correspond to a shortening at low frequencies by a factor of 0.61—this is consistent with my own measurements. The technique I used to measure the effects of compensation was to compare resonance frequencies of tubes in which the bend was located close to a pressure maximum with those in which the bend was located near a pressure minimum. These techniques are described in detail by Coltman.[3] Experiments were done with plastic tubes, though the results are applicable to metal.

III. COMPENSATION

Compensation to restore the characteristic impedance to that of the straight tube can be obtained either by increasing the inertance to match the compliance, decreasing the compliance to match the inertance, or both. The acoustic length of the section may thereby be shortened, which must be accounted for in the instrument design.

In Fig. 1(a), the first method, the inertance has been increased by making a saw cut of depth h at 45 deg in the inner corner and inserting a thin card (in the experiment a plastic card of thickness 0.04 ID was used) and cementing it
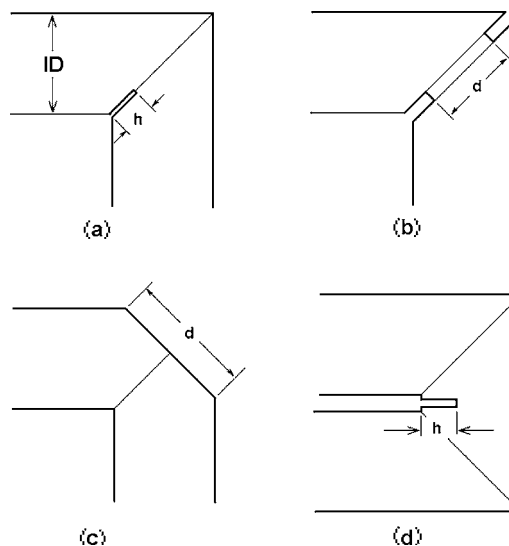
FIG. 1. Methods to compensate miter bends. Only inside dimensions shown.

[a]Electronic mail: coltmanjw@verizon.net

in place. The proper value to achieve zero reflection is $h = 0.35$ ID. The edge of the card is preferably rounded to reduce turbulence. The acoustical length of the section is essentially equal to the physical length.

An alternative method is shown in Fig. 1(b). Here a disk of thickness 0.1 ID contains a round hole of diameter $d = 0.87$ ID and is cemented between the tubes as they are joined. Rounding of the edges to reduce turbulence is recommended. The compensation will depend somewhat on the thickness of the disk, but since the hole is not far in diameter from that of the tubing, the effect will not be large; a variation from 0.05 to 0.15 ID probably would not change things much. The acoustic length of this section is close to the distance measured along the centerlines.

In Fig. 1(c), the compliance has been reduced, and the inertance somewhat increased, by beveling the miter and closing the hole with a plate. The dimension d is that of the long axis of the near-elliptical hole created by beveling the miter; its proper value is $d = 1.26$ ID. The acoustical distance around the bend is shortened by 0.32 ID.

A 180-deg bend or fold can be produced by two successive 90-deg bends, the tightest fold being that when the outside walls of the tubes are in contact, as shown in Fig. 1(d). I found that the acoustic shortening due to this assembly was not much different than the sum for two distant 90-deg bends, so wall thickness of a tight bend can be ignored as a variable. Here compensation can be provided by a segment of a disk inserted into the center of the bridging tube. The proper segment was found to have a height that depended on its thickness t, the proper value being $h = 0.49$ ID $- 0.5t$. The acoustic length of this assembly will be shortened by about $0.4t$.

Finally, the double bend can be compensated by beveling each miter as was done in Fig. 1(c). This is probably the most esthetically pleasing arrangement for a fold. Because of the interaction of the close bends, the proper value for d is not the same as in Fig. 1(b). I found it to be $d = 1.33$ ID, and the acoustic shortening of the assembly to be 0.69 ID.

IV. CONCLUSION

Compensation of miter bends can be readily accomplished. The numbers given here are estimated to be applicable for practical purposes to wavelengths greater than four times the inside diameter of the tubing.

[1] C. J. Nederveen, "Influence of a toroidal bend on wind instrument tuning," J. Acoust. Soc. Am. **104**(3), 1616–1620 (1998).
[2] S. Dequand, S. J. Hulshoff, A. Ayregan, J. Huijnene, R. ter Riet, L. J. van Lier, and A. Hirschberg, "Acoustics of 90 degree sharp bends. Part 1: Low-frequency response," Acta Acust. **89**, 1025–1037 (2003).
[3] J. W. Coltman, "Acoustic properties of miter bends," http://ccrma.stanford.edu/marl/Coltman/Papers.html, item 1.44, December 2006, 188kB, PDF (archived), viewed 4/6/2007.

Acoustic cues available for ranging by humpback whales (L)

Eduardo Mercado III,[a] Jennifer N. Schneider, Sean R. Green, Chong Wang,
Rachael D. Rubin, and Patchouly N. Banks
Department of Psychology, University at Buffalo, The State University of New York, Buffalo, New York
14260

Field measurements of sound propagation in a humpback whale habitat were collected to identify cues that a humpback whale might use to estimate its distance from sound sources. The data show that spectral cues are sufficient for estimating the relative distance a sound has traveled in such environments, and that several other cues may also provide useful information. It is suggested that listening humpback whales may use multiple cues in parallel to determine the range to singing whales. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717495]

I. INTRODUCTION

Sound waves propagating through an environment are affected by the physical features of that environment. A common result of such physical interactions is sound degradation. For example, constructive and destructive interference can produce complex fluctuations in both the amplitude and frequency of sound waves. In animal communication research, such distortion is generally viewed as a problem that organisms must solve in order to recognize and appropriately interpret acoustic signals that have traveled long distances. In terms of localizing sounds, however, such "degradation" can be an important source of information about both the environment and the sound-producing organisms within the environment. Numerous birds and mammals (including humans) use environmental distortion to judge the distance that a sound has traveled in a process known as ranging.[1–4]

Whereas ranging is well established in birds and primates, it has yet to be demonstrated in most other species of animals. Recently, simulations of underwater acoustic propagation led to the surprising prediction that humpback whales (*Megaptera novaeangliae*) might be able to range vocalizing conspecifics over several kilometers using spectral degradation cues.[5] Ranging has yet to be demonstrated in any cetacean, but their sophisticated auditory capabilities and observed behavioral interactions strongly suggest that they should be able to judge the distance to a sound source.

The goal of this study was to examine how humpback whale-like sounds degrade as they propagate through a near-shore environment frequented by humpback whales, in order to identify acoustic features that change systematically with distance.[4,6–10] Reliable patterns of degradation provide useful cues that humpback whales might use for ranging, whereas unpredictable distortions do not.

II. METHODS

Sound transmission experiments were conducted off the northwestern coast of Puerto Rico where humpback whales are consistently seen and heard,[11–13] at water depths that whales are known to sing (50–550 m) (Fig. 1(A)). Experiments were performed during March 2006, between 8:30 am and 12:30 pm; water temperature was approximately 25 °C throughout the first 100 m of the water column. The current experiments were conducted when no whales were visibly present, and when song amplitude suggested that no audible singers were nearby.

Narrowband tones and broadband sounds, similar in frequency content (100–22,000 Hz) and duration (2 s) to components of humpback whale songs, were broadcast from an underwater speaker, and recorded at locations both near to and remote from the speaker (Fig. 1(B)); each sound was broadcast 20 times from each position. Figure 1(C) illustrates the time-frequency structure of one of the broadband sounds (a hyperbolic, frequency-modulated tonal sweep) recorded about 2 m from the underwater speaker, and Fig. 1(D) shows how the acoustic features of this sound changed after it had propagated 800 m. The changes apparent in Fig. 1(D) are representative of the basic effects of long-range propagation in shallow water—signal amplitude is attenuated, energy at particular frequencies is decreased, and reverberations increase.

Thirteen different tonal and frequency-modulated sounds, varying in waveform shape and spectral bandwidth (but normalized with respect to peak waveform amplitude), were generated in .wav format using MATLAB Version 6.5. These sounds included broadband white noise, pure tones, frequency-modulated sweeps, sawtooth waves, and pulse trains (repetition rate=0.3 Hz for 1 min); .wav files were converted into .mp3 format using Sonic Stage Software (Ver. 3.0) and uploaded as 13 tracks to a Sony MZ-M100 Hi-MD Recorder. Note that inharmonic partials evident in Figs. 1(C) and 1(D) are likely the result of distortions in the original signal introduced by the nonlinear (mp3) format used in sound playbacks. The recorder was connected to a Speco PAT-20 20 W, 12 V dc, Mobile Amplifier with Microphone (through the auxiliary input), which connected to a Lubell LL916 Underwater Speaker with an AC202 (50 W, 8 ohm, 20 V rms) transformer box, to produce the sounds. Low source levels (<150 dB re 1 μPa) were used to insure that playbacks would not adversely impact any undetected

---
[a] Author to whom correspondence should be addressed. Electronic mail: emiii@buffalo.edu
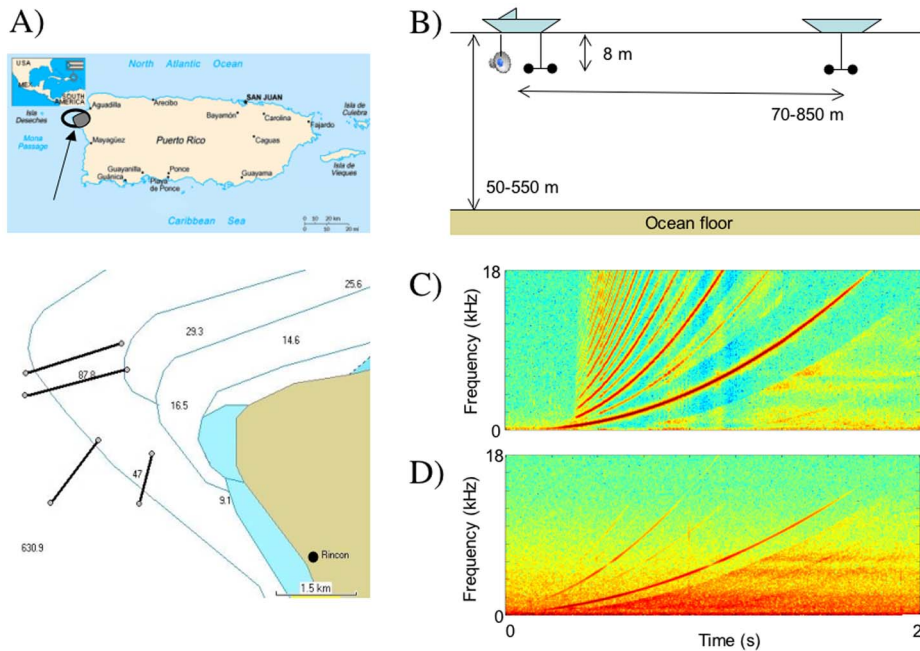
FIG. 1. (Color online) Sound transmission experimental setup. (A) Map of the northwestern coast of Puerto Rico showing the locations of the direct paths (lines with circles) of sound transmission between the underwater speaker and recording hydrophones. Circles closest to land represent the start point of sound source. Drift occurred in the direction of transmission. Curved lines are depth indicators provided in BlueChart Americas v7.5. (B) The experimental configuration emulates the situation faced by a listening whale near the ocean surface. (C) Spectrogram (FFT size=512 samples, hanning window, 50% overlap) of a broadband frequency modulated sweep recorded near the transmitting boat. (D) Spectrogram of this same sound after having travelled 800 m, as recorded from the second boat, shows attenuation of the lowest and highest frequencies, as well as spectral notches indicative of destructive interference.

whales in the area. The speaker was lowered to an approximate depth of 7 m from the side of a small boat. All 13 tracks on the mini disk recorder were played sequentially.

Broadcast sounds were recorded using two hydrophone arrays at a sampling rate of 44.1 kHz. Each array consisted of a pair of hydrophones (Cetacean Research C10; 0.25–25 kHz flat frequency range, ±3 dB), connected to a pair of pre-amplifiers (Cetacean Research Model SS03), which fed into a digital recorder (Sony MD Walkman Mz-NH900, recording in .wav format). The paired hydrophones were fixed 0.3 m apart from each other; this distance corresponds to the spacing between an adult humpback whale's ears. The hydrophones were positioned approximately 8 m below the surface. One system was suspended from the side of the boat, approximately 2 m from the sound source, and the other system was suspended from a kayak at ten distances ranging between 70–850 meters from the speaker (approximately 75 m spacing; best fitting line=74.9x+71.5; $R^2=0.99$). Positions of the boat and kayak were determined using a GPS receiver (~3–5 m horizontal accuracy). The study area ranged between N 18°22.491 to N 18°20.961 and W 67°18.703 and W 67°17.068.

Initial analyses were conducted to determine how signal degradation varied as a function of distance traveled, sound type, and the water depth where the receiving system was located. First, automated measures of the maximum peak-to-peak amplitude were collected from all recorded wave forms and correlated with measures of distance and water depth. The five sound types that produced the highest received levels (i.e., highest signal-to-noise ratios, S/N, at the farthest recording distance) were selected for detailed analysis. These consisted of 500 and 1000 Hz tones, linear and quadratic frequency sweeps, and a 50 Hz sawtooth wave. The remaining eight sound types showed similar patterns of degradation to this subset at shorter distances. Three representative measures of signal degradation comparable to those used in past propagation research[6] were then computed for these five

sound types: S/N, attenuation, and blur ratio. S/N was computed from the energy of the received sound and the energy of the background noise sampled from intervals between sounds, attenuation was computed relative to the received level recorded at the shortest range, and blur ratio was computed by comparing the energy of the difference amplitude function (an attenuated version of the envelope of the source wave form) to the energy of the received amplitude function (see Ref. 7 for further details). Raw attenuation measures were used rather than the excess attenuation measures used in studies of bird song degradation due to the complex physics of sound propagation in shallow water environments. For 10% of recordings, the transmitted signal was lower in amplitude than the noise background. Measures from these recordings were not included in statistical analyses, except in the case of correlation calculations, for which they were set to zero. The data for each of these three measures were subjected to analysis of covariance (ANCOVA), and the extent to which each measure related to propagation distance and water depth was examined by linear regression. The relationship between the three measures was tested by calculating correlation coefficients for each pair of measures and testing for significance. Next, power spectral density functions were estimated for each received sound ($n=1000$) using MATLAB [fast Fourier transform (FFT)=16,384 samples, 44.1 kHz sampling rate]. The resulting vectors were then truncated to reduce the maximum frequency to 18 kHz.

III. RESULTS

The amplitude of received sounds was negatively correlated with distance as a result of transmission loss ($r=-0.35$, $n=2600$, $P<0.0001$). Amplitude changes in broadband sounds were more strongly correlated with range ($r=-0.44$, $n=1800$, $P<0.0001$) than changes in narrowband sounds ($r=-0.07$, $n=800$, $P=0.06$). Amplitude changes were also correlated with water depth, but to a lesser extent and in the opposite direction ($r=0.17$, $n=2600$, $P<0.0001$).
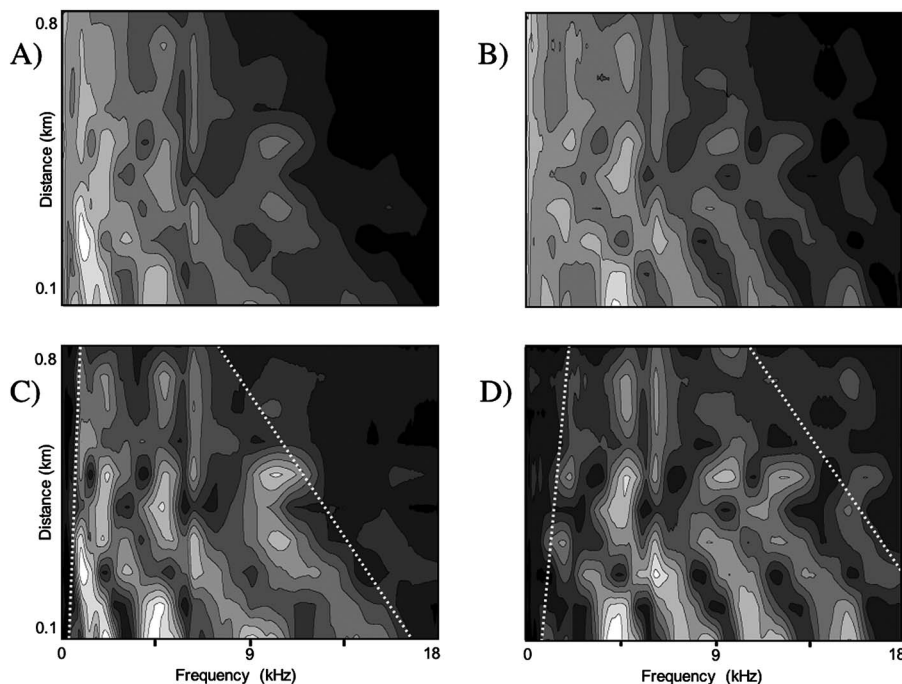
FIG. 2. Spectral degradation as a function of distance. (A) Contour plot of the spectral peaks (white-light gray) and valleys (black) of the hyperbolic frequency sweep shown in Fig. 1(C), recorded at ten different distances reveals a complex pattern of propagation in which spectral peaks can decrease with distance, but then reappear at farther ranges. Additionally, some higher frequencies show less attenuation than adjacent lower frequencies. (B) Contour plot of distance-dependent degradation of a linear frequency sweep spanning the same range of frequencies as the hyperbolic sweep shows similar patterns. (C) Dividing the received spectra of the hyperbolic sweep by the spectra of ambient noise reveals that spectral peaks and notches are shifting to lower frequencies with distance. (D) Normalizing the linear sweep spectra to noise also reveals that both the highest and lowest frequencies are attenuating faster than mid-range frequencies. Dotted white lines show how this effect gradually narrows the bandwidth of a broadband sound as a function of distance propagated (end points of the lines correspond to boundaries of level contours, shown in gray, at the nearest and farthest distances).

S/N, attenuation, and blur ratio were correlated for all sound types. As expected, S/N decreased with increased attenuation ($r=-0.33$, $n=1000$, $P<0.001$), and with increased blur ratio ($r=-0.59$, $n=1000$, $P<0.001$); attenuation increased as blur ratio increased ($r=0.25$, $n=1000$, $P<0.001$). Signal degradation varied with distance as a function of sound type: S/N (ANOVA$_{Distance \ x \ Type}$ $F_{36,847}=10.336$, $P<0.001$), attenuation (ANOVA$_{Distance \ x \ Type}$ $F_{32,855}=89.25$, $P<0.001$), and blur ratio (ANOVA$_{Distance \ x \ Type}$ $F_{32,855}=85.82$, $P<0.001$). S/N decreased with distance for all sound types. Attenuation increased with distance for both tones and quadratic sweeps, but not linear sweeps or saw-tooth waves. Blur ratio increased with distance for all sounds except sawtooth waves. To confirm that these findings were not simply resulting from variability in water depth, the same analysis was performed using depth as a factor. S/N decreased with depth for tones only. Attenuation increased with depth for both types of sweeps and sawtooth waves. Blur ratio increased with depth for linear sweeps and sawtooth waves.

Changes in the spectra of broadband, frequency-modulated tonal sweeps as a function of range were analyzed to determine whether lower frequencies within sounds propagated farther than higher frequencies (Fig. 2(A) and 2(B)). For these sounds, frequencies above 13 kHz attenuated more rapidly than lower frequencies as a function of distance ($t=-17.1$, $df=6686$, $P<0.0001$, *one-tailed*). It was not the case, however, that the lowest frequencies consistently propagated the best. Rather, there appeared to be particular midrange frequencies that propagated better than other frequencies,

and predictable patterns of alternating better and worse propagating frequencies, consistent with the predictions of prior simulations.[5] The attenuation of low frequency components is more clearly evident when received signals are normalized relative to ambient noise (Fig. 2(C) and 2(D)). In general, the greater attenuation of the highest and lowest frequencies within a sound produced a systematic narrowing of the signal bandwidth. The distribution of spectral peaks and notches within this band also differentiated particular distances.

IV. DISCUSSION AND CONCLUSIONS

The finding that midrange frequencies often propagate better in shallow water environments than lower frequencies provides a possible explanation for why humpback whales produce higher frequency sounds than most other baleen whales. Humpback whales can produce sounds with fundamental frequencies of 1–4 kHz, which is significantly higher than is observed in vocalizations produced by many smaller mammals, including most humans. The current results suggest that the multi-octave sounds that humpback whales produce when singing maximize their transmission range, and provide cues for distance estimation that would not be available if they used the infrasonic tonal sounds used by many other baleen whales.

These analyses revealed a simple cue for distance estimation that is relatively resistant to ocean noise conditions, as well as fluctuations in amplitude. Specifically, spectral bandwidth narrowed systematically as a function of distance. Although changes in amplitude were also correlated with

propagation distance, amplitude cues can only provide gross information about source location, and require some knowledge of source features to be useful. Both amplitude and spectral bandwidth cues are more useful for ranging broadband sounds than narrowband sounds. Peaks and notches within received spectra may provide more precise localization cues that are less dependent on source levels and bandwidth. This effect is analogous to direction-dependent filtering by an animal's head,[14,15] except that in terrestrial auditory reception, spectral peaks and notches can provide information about source azimuth and elevation, whereas in shallow water aquatic environments, similar spectral cues can provide information about source distance (and possibly bearing in certain conditions[16]). Whales likely make use of all of these acoustic cues, and others such as the ratio of direct to reverberant sound, when estimating source distance.

The role that internal representations of source signals play in localizing sounds is widely acknowledged for some animals (echolocating bats[17] and humans[18]), debated for others (songbirds[19]), and ignored for most, including cetaceans. It is especially clear in the case of singing humpback whales, however, that individuals maintain good representations of sounds as they would be heard near the source, because most or all of the singers within a particular region and time period use highly similar sound repertoires that they collectively change over time.[20–23] There is increasing awareness that auditory spatial cues can contribute to the social interactions of cetaceans;[24,25] however the role that experience with particular sounds or environments plays in this process remains poorly understood. The current data suggest that, although knowledge of source features can facilitate ranging by humpback whales, it is not a necessity. As long as spectral energy within produced sounds consistently spans a wide bandwidth, accurate range estimation should be possible without knowledge of specific source features.

Sound localization has previously been identified as a driving force in the evolution of mammalian hearing,[26] but the focus has traditionally been on azimuthal resolution rather than on distance estimation. Ranging capacity has, however, previously been proposed to be critical in the evolution of song learning and perception by birds.[27] The current findings suggest that similar constraints may also have guided the evolution of sound perception and production by humpback whales. Armed with the knowledge of ranging cues that are available to whales, researchers can better assess the extent to which different cetacean species use such cues. This knowledge may also be useful in understanding how human-made noises interfere with whales' abilities to localize conspecifics and avoid collisions with boats.

ACKNOWLEDGMENTS

[1] P. McGregor, "Sound cues to distance: The perception of range," in *Perception and Motor Control in Birds*, edited by M. N. O. Davies and P. R. Green (Springer-Verlag, Berlin, 1994), pp. 74–94.

[2] M. Naguib and R. H. Wiley, "Estimating the distance to a source of sound: Mechanisms and adaptations for long-range communication," Anim. Behav. **62**, 825–837 (2001).

[3] P. M. Waser, "Sound localization by monkeys: A field experiment," Behav. Ecol. Sociobiol. **2**, 427–431 (1977).

[4] D. G. Richards and R. H. Wiley, "Reverberations and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication," Am. Nat. **115**, 381–399 (1980).

[5] E. Mercado III and L. N. Frazer, "Environmental constraints on sound transmission by humpback whales," J. Acoust. Soc. Am. **106**, 3004–3016 (1999).

[6] T. Dabelsteen, O. N. Larsen, and S. B. Pedersen, "Habitat-induced degradation of sound signals: Quantifying effects of communication sounds and bird location on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song," J. Acoust. Soc. Am. **93**, 2206–2220 (1993).

[7] J. Holland, T. Dabelsteen, S. B. Pedersen, and O. N. Larsen, "Degradation of wren *Troglodytes troglodytes* song: Implications for information transfer and ranging," J. Acoust. Soc. Am. **103**, 2154–2166 (1998).

[8] E. S. Morton, "Ecological sources of selection on avian sounds," Am. Nat. **109**, 17–34 (1975).

[9] S. L. Gish and E. S. Morton, "Structural adaptations to local habitat acoustics in Carolina wren songs," Z. Tierpsychol **56**, 74–84 (1981).

[10] K. Marten, D. Quine, and P. Marler, "Sound transmission and its significance for animal vocalization. II. Tropical forest habitats," Behav. Ecol. Sociobiol. **2**, 291–302 (1977).

[11] I. M. Sanders, J. C. Barrios-Santiago, and R. S. Appledoorn, "Distribution and relative abundance of humpback whales off Western Puerto Rico during 1995–1997," Caribb. J. Sci. **41**, 101–107 (2005).

[12] S. L. Swartz, A. Martinez, J. Stamates, C. Burks, and A. A. Mignucci-Giannoni, "Acoustic and visual survey of cetaceans in waters of Puerto Rico and the Virgin Islands: February–March 2001," NOAA Technical Memorandum Report No. NMFS-SEFSC-463 (2001).

[13] H. E. Winn and L. K. Winn, "The song of the humpback whale *Megaptera novaeangliae* in the West Indies," Mar. Biol. (Berlin) **47**, 97–114 (1978).

[14] M. Aytekin, E. Grassi, M. Sahota, and C. F. Moss, "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," J. Acoust. Soc. Am. **116**, 3594–3605 (2004).

[15] A. Kulkarni and H. S. Colburn, "Role of spectral detail in sound-source localization," Nature (London) **396**, 747–749 (1998).

[16] L. N. Frazer and P. I. Pecholcs, "Single hydrophone localization," J. Acoust. Soc. Am. **88**, 995–1002 (1990).

[17] J. A. Simmons, M. J. Ferragamo, and C. F. Moss, "Echo-delay resolution in sonar images of the big brown bat, *Eptesicus fuscus*," Proc. Natl. Acad. Sci. U.S.A. **95**, 12647–12652 (1998).

[18] P. McGregor, A. G. Horn, and M. A. Todd, "Are familiar sounds ranged more accurately?," Percept. Mot. Skills **61** (3 Pt 2), 1082 (1985).

[19] E. S. Morton, "Degradation and signal ranging in birds: Memory matters," Behav. Ecol. Sociobiol. **42**, 135–137 (1998).

[20] L. N. Guinee, K. Chu, and E. M. Dorsey, "Changes over time in the songs of known individual humpback whales (*Megaptera novaeangliae*)," in *Communication and Behavior of Whales*, edited by R. Payne (Westview, Boulder, CO, 1983).

[21] K. Payne and R. Payne, "Large scale changes over 19 years in songs of humpback whales in Bermuda," Z. Tierpsychol **68**, 89–114 (1985).

[22] M. J. Noad, D. H. Cato, M. M. Bryden, M. N. Jenner, and K. C. S. Jenner, "Cultural revolution in whale songs," Nature **408**, 537 (2000).

[23] E. Mercado, L. M. Herman, and A. A. Pack, "Song copying by humpback whales: Themes and variations," Anim. Cog. **8**, 93–102 (2005).

[24] P. J. O. Miller, "Mixed-directionality of killer whale stereotyped calls: A direction of movement cue?," Anim. Behav. **52**, 262–270 (2002).

[25] M. O. Lammers and W. W. L. Au, "Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?," Marine Mammal Sci. **19**, 249–264 (2003).

[26] R. S. Heffner and H. E. Heffner, "Evolution of sound localization in mammals," in *The Evolutionary Biology of Hearing*, edited by D. W. Webster, R. R. Fay, and A. N. Popper (Springer-Verlag, New York, 1992), pp. 691–715.

[27] E. S. Morton, "Why songbirds learn songs: An arms race over ranging?," Poul. Avian Biol. Rev. **7**, 65–71 (1996).

An efficient high-order algorithm for acoustic scattering from penetrable thin structures in three dimensions

Akash Anand[a]
Applied and Computational Mathematics, California Institute of Technology, Pasadena, California 91125

Fernando Reitich[b]
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

This paper presents a high-order accelerated algorithm for the solution of the integral-equation formulation of volumetric scattering problems. The scheme is particularly well suited to the analysis of "thin" structures as they arise in certain applications (e.g., material coatings); in addition, it is also designed to be used in conjunction with existing low-order FFT-based codes to upgrade their order of accuracy through a suitable treatment of material interfaces. The high-order convergence of the new procedure is attained through a combination of changes of parametric variables (to resolve the singularities of the Green function) and "partitions of unity" (to allow for a simple implementation of spectrally accurate quadratures away from singular points). Accelerated evaluations of the interaction between degrees of freedom, on the other hand, are accomplished by incorporating (two-face) equivalent source approximations on Cartesian grids. A detailed account of the main algorithmic components of the scheme are presented, together with a brief review of the corresponding error and performance analyses which are exemplified with a variety of numerical results. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2714919]

I. INTRODUCTION

The evaluation of scattering returns from large three-dimensional penetrable scatters continues to constitute one of the most challenging problems in computational science. The main difficulties that are present in the problem arise from the need to accurately describe and manipulate highly oscillatory functions. Although a large body of work has been devoted to the subject, scattering problems still cannot be considered completely solved, particularly from a computational perspective, and they remain the subject matter of much ongoing research. A main reason for this lies in the dependence of the mathematical formulation and computational complexities on the oscillations of the incident radiation and on the geometrical and acoustical (or other physical and/or material properties, as the case might be) specifics of the scatterer. The latter include, for instance, their geometrical shapes and material densities, velocities, absorption, and penetrability characteristics, to name but a few.

This diversity of parameters and the different nature of the difficulties associated with particular instances make the case for specialized, application specific algorithms. One such important instance is that of scattering by three-dimensional "thin" scatterers. Thin (thickness up to the order of the wavelength of the incoming radiation) scatterers are used, for instance, to model coatings. Clearly, these have a wide range of practical applications, both for civilian and military purposes (e.g., decoy detection, stealth technology, etc). In this paper, we present a new, efficient and high-order

solver for the prediction of the acoustic response of such thin (and possibly inhomogeneous) structures, that is based on the integral equation formulation of the scattering problem. As we explain in the following, and beyond their intrinsic interest and applicability, our new schemes can also be viewed as providing a mechanism to upgrade current low-order (e.g., FFT-based) solvers to fast high-order schemes in the presence of material discontinuities.

Most existing algorithms for computing solutions to volumetric scattering problems can be broadly classified into three classes, depending on the underlying mathematical formulation of the problem: (i) Differential equation methods; (ii) variational methods; and (iii) integral equation methods. The methods in category (i) are based on direct discretization of the differential equation [e.g., Eq. (1)] and they include, most notably, finite difference schemes (e.g., FDTD techniques[1–3]). Those in category (ii), in contrast, work with the variational form of the equation, such as in all variants of the finite element (FE) method.[4–6] In both cases, however, a central challenge facing these numerical procedures in connection with scattering applications relates to the need to truncate the (infinite) physical domain. This issue has received, and continues to receive, significant attention in the literature as the imposition of approximate absorbing or transparent boundary conditions, or the use of perfectly matched layers (PML), can still yield, in some cases, unacceptable results. On the other hand, even when an approximate treatment of the truncation boundary/region does not pollute the numerical solution (e.g., when this is effected at a considerable distance from the scatterers), other sources of error may produce unwanted results. In this regards, perhaps the most prominent source can be attributed to the dispersive

---

[a]Electronic mail: anand@acm.caltech.edu
[b]Electronic mail: reitich@math.umn.edu

characteristics of standard (low-order) methods based on differential and variational formulations. A classical remedy resorts to smaller grid spacing, with a consequent increase in computational effort.

A number of more sophisticated solutions to the low accuracy of standard FDTD and/or FE methodologies, however, have been developed. Generically, these algorithms are based on the use of higher-order discretization schemes. Most notable among these are so-called "spectral methods,"[7–11] based on the use of global spatial basis functions, that provide spectral accuracy in the approximation of smooth functions. In the time domain, different methods within this category handle temporal derivatives differently.

For example, the pseudospectral time-domain (PSTD) methods[7,8] use conventional differencing schemes, e.g., leapfrog, to evolve the solution in time. The "k-space method," originally proposed by Bojarski[12,13] and further developed by Mast et al.,[10] on the other hand, uses a nonstandard finite difference approach for temporal integration, which is exact for the homogeneous case; spectral approximations in time have also been used.[14]

While extremely efficient and accurate within their domain of applicability, the implementation of spectral methodologies is, however, nontrivial, and their convergence properties, specifically for those based on the more efficient Fourier spectral approximations, are dependent on the underlying material properties. Indeed, for instance, when applied in a straightforward manner, these techniques suffer from so-called "wraparound effects," caused by the lack of periodicity in the problem. As has been recently shown in the context of the PSTD and k-space approaches,[9,11] however, these effects can be significantly mollified by employing a PML. In this manner, and for practical purposes, the resulting schemes can be considered to converge with higher orders, at least in simulations that entail smooth scattering media. The extension to scattering configurations that present material discontinuities, on the other hand, is substantially impeded by Gibbs-type phenomena that are inherent to spectral approximations. The modifications that have been proposed to date to alleviate these effects are not as satisfactory: Those based on filtering[10,11] or NUFFTs[15–17] clearly do not retain high-order convergence characteristics, while those that use postprocessing procedures (e.g., Gegenbauer reconstruction[18]) are of limited applicability[19,20] and, moreover, they can adversely affect the computational cost.

As we mentioned, the work we present here relates to a new methodology for the solution of the (volumetric) integral equation formulation of the scattering problem that provides a high-order treatment of material discontinuities. Our approach, therefore, falls within the above-mentioned category (iii) and, as such, it obviates the need for the explicit imposition of radiation conditions (as these are encoded in the Green's function). Within this class of methods, perhaps the best known algorithms are those related to the CGFFT,[21,22] which uses an iterative procedure (conjugate gradients, in this case) to solve the (convolution) integral equation; at each iteration the convolution is effected with fast Fourier transforms which, for a number N of unknowns, entails a cost of $O(N \log N)$ [instead of the $O(N^2)$ cost of direct evaluation]. Again here, this solution procedure leads to very efficient numerics and is highly accurate for smoothly varying media, but its convergence degrades significantly in the presence of material discontinuities.[17,23] In fact, to our knowledge, only limited attempts have been made toward the development of efficient higher-order integral-equation solvers for volumetric scattering applications. A notable exception is the work of Zhu and Gedney,[24,25] based on the "discontinuous FFT" of Fan and Liu.[23] This latter scheme is based on the accurate evaluation of Fourier coefficients of discontinuous functions (through Gaussian quadratures and careful interpolation from and to equispaced grids) and, while it can be shown that this improves on the convergence of the CGFFT (from first- to second-order accurate, in fact), it fails to address the Gibbs phenomenon which arises as the Fourier series is summed. A simpler and more efficient approach, based on a suitable decomposition of the Green's function, and also leading to second-order convergence for discontinuous scatterers, has been recently proposed in Ref. 26.

The rest of the paper is organized as follows. In Sec. II, we precisely define the concept of a "thin structure" and we review the mathematical formulation of the corresponding volumetric scattering problem, based on the solution of the Helmholtz equation. The main algorithmic components of our approach are detailed in Sec. III. In Sec. IV, an analysis of the accuracy and efficiency of the numerical scheme is presented. This is followed, in Sec. V, by a presentation of the results from a variety of numerical experiments that exemplify the properties of the procedure. We present, in Sec. VI, a brief discussion on the use of our thin volume integrator in conjunction with existing low-order FFT-based methods to upgrade their order of accuracy in the presence of material discontinuity. Finally, our conclusions are summarized in Sec. VII.

II. PRELIMINARIES

A. Mathematical formulation

As mentioned earlier, we shall consider the Helmholtz model of acoustic scattering by a bounded, inhomogeneous medium Ω of refractive index $n(\mathbf{x})$. Within this context, given an incident field u^i, its interaction with the scatterer produces a scattered field u^s and the total field $u = u^i + u^s$, satisfies the Helmholtz equation

$$\Delta u(\mathbf{x}) + \kappa^2 n^2(\mathbf{x}) u(\mathbf{x}) = 0, \quad \mathbf{x} \in \mathbb{R}^3. \tag{1}$$

In addition, the Sommerfeld radiation condition is imposed on u^s to enforce its outgoing character. Equivalently,[27] the total field u also satisfies the Lippmann-Schwinger integral equation

$$u(\mathbf{x}) + \kappa^2 K[u](\mathbf{x}) = u^i(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^3, \tag{2}$$

where

$$K[u](\mathbf{x}) = \int_\Omega G(\mathbf{x}, \mathbf{x}') m(\mathbf{x}') u(\mathbf{x}') d\mathbf{x}', \tag{3}$$
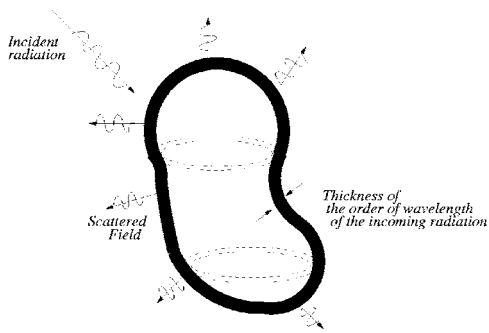
(a) Scattering by "thin" structures
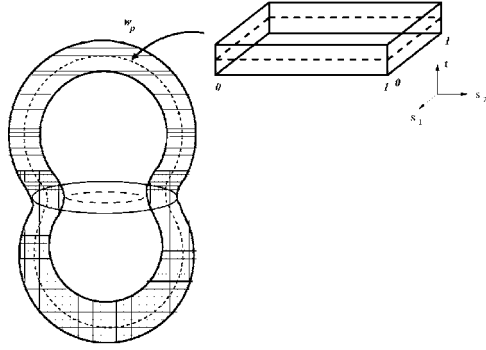
(b) A Thin Structure

FIG. 1. Description of thin structures; overlapping patches and corresponding parametric spaces are constructed to cover the scatterer of thickness of the order of the wavelength of incoming radiation.

$$G(\mathbf{x},\mathbf{x}') = \frac{\exp(\iota\kappa|\mathbf{x}-\mathbf{x}'|)}{4\pi|\mathbf{x}-\mathbf{x}'|}$$

is the Green's function for the Helmholtz equation in three dimensions and $m = 1 - n^2$. Note that $m \equiv 0$ outside of Ω and also that the choice of Green's function guarantees that u^s is outgoing. The solvability of the integral equation (2) has been thoroughly investigated, and positive results have been established for rather general contrast functions m; see, e.g., Refs. 27–30

As we said, our initial developments will be geared toward solvers for thin structures which we formally define next.

B. Thin structures

A thin structure Ω (see Fig. 1) is described by a finite collection of patches $\mathcal{P} = \{\mathcal{P}^p : p = 1, \dots, P\}$ that cover the entire structure, i.e., $\Omega = \cup_{p=1}^{P} \mathcal{P}^p$, where each patch is parameterized by a smooth invertible map $\mathbf{x}^p : (0,1)^3 \to \mathcal{P}^p$,

$$\mathbf{x}^p = \mathbf{x}^p(\mathbf{s},t), \quad (\mathbf{s},t) = (s_1, s_2, t) \in (0,1)^3.$$

By convention, the two-dimensional variable $\mathbf{s}$ shall correspond to "parallel" coordinates while the variable t will describe the structure in the "transverse" direction. *We say that the structure Ω is thin if the thickness*

$$\sup_{\mathbf{s} \in (0,1)^2} |\mathbf{x}^p(\mathbf{s},1) - \mathbf{x}^p(\mathbf{s},0)| \lesssim \lambda \tag{4}$$

for all $p = 1, \dots, P$. Here, $\lambda = 2\pi/\kappa$ is the wavelength of the incident radiation.

III. AN EFFICIENT HIGH-ORDER SOLVER

Naturally, the possibility of producing fast and accurate solutions depends on our ability to evaluate the integral in Eq. (3) accurately and efficiently. Clearly, $G(\mathbf{x},\mathbf{x}')$ changes rapidly when the source point $\mathbf{x}'$ is close to or, in other words, *adjacent* to the target point $\mathbf{x}$ and is singular when $\mathbf{x} = \mathbf{x}'$. Thus, the first difficulty relates to the design of high order quadratures for $K[u](\mathbf{x})$ in the vicinity of the target point $\mathbf{x}$. More precisely, for a neighborhood $\mathcal{N}_\mathbf{x}$ of $\mathbf{x}$, we can separate the integral in Eq. (3) into

$$K[u](\mathbf{x}) = \int_{\mathcal{N}_\mathbf{x}} G(\mathbf{x},\mathbf{x}')m(\mathbf{x}')u(\mathbf{x}')d\mathbf{x}'$$
$$+ \int_{\Omega\backslash\mathcal{N}_\mathbf{x}} G(\mathbf{x},\mathbf{x}')m(\mathbf{x}')u(\mathbf{x}')d\mathbf{x}' \tag{5}$$

corresponding to "adjacent" and "nonadjacent" interactions, respectively; the calculation of the (singular) adjacent interactions then necessitates the design of a specialized quadrature rule (Sec. III A). Once these *adjacent interactions* have been computed, there will only remain to compute nonadjacent interactions

$$\int_{\Omega\backslash\mathcal{N}_\mathbf{x}} G(\mathbf{x},\mathbf{x}')m(\mathbf{x}')u(\mathbf{x}')d\mathbf{x}'.$$

The accurate evaluation of these though does not pose a major difficulty in the design of approximate quadratures as the integrand is regular. However their number [entailing $\mathcal{O}(N)$ operations for each of N target points] demands the design of a suitable "accelerator" (Sec. III B). In summary, the goal is to design an algorithm that accurately evaluates the singular *adjacent interactions* without compromising speed and that can, at the same time, efficiently evaluate the voluminous number of *nonadjacent interactions*.

A. Adjacent interactions

The singular nature of the integral kernel $G(\mathbf{x},\mathbf{x}')$ causes the main difficulty in high-order evaluation of *adjacent interactions*. Even though the use of the well-known strategy of "singularity subtraction" does give rise to bounded integrands, an application of classical high-order quadratures to these still fails to provide high-order results, since the higher order derivatives of the integrand are unbounded. Thus, a specialized quadrature rule must be developed and used to achieve high-order accuracy.

To present the method, consider a partition of unity subordinated to the covering $\mathcal{P}$, that is a collection of functions

$$w^p(\mathbf{x}), \quad p = 1, \dots, P, \tag{6}$$

such that

(1) w^p is defined, smooth, and non-negative in Ω, and it vanishes outside $\mathcal{P}^p$, and
(2) $\Sigma_{p=1}^{P} w^p = 1$ throughout Ω.

Then, we can write

$$K[u](\mathbf{x}) = \sum_{p=1}^{P} K_p[u](\mathbf{x}),$$

where

$$K_p[u](\mathbf{x}) = \int_0^1 \left(\int_{[0,1]^2} G(\mathbf{x},\mathbf{x}^p(\mathbf{s}',t')) \phi_u^p(\mathbf{s}',t') d\mathbf{s}' \right) dt' \tag{7}$$

and

$$\phi_u^p(\mathbf{s},t) = m(\mathbf{x}^p(\mathbf{s},t)) u(\mathbf{x}^p(\mathbf{s},t)) w^p(\mathbf{x}^p(\mathbf{s},t)) J^p(\mathbf{s},t).$$

Here, J^p is the Jacobian of the transformation $\mathbf{x}^p$.

In the case when $\mathbf{x}$ does not lie on the integration patch $\mathcal{P}^p$, the integrand in Eq. (7) is smooth and as a result can be integrated to high order. The fact that the integrands in the two-dimensional "parallel" integrals

$$\int_{[0,1]^2} G(\mathbf{x},\mathbf{x}^p(\mathbf{s}',t')) \phi_u^p(\mathbf{s}',t') d\mathbf{s}'$$

vanish to high order at the boundary of the integration domain $[0,1]^2$ (due to the presence of w^p as a factor) suggests the use of trapezoidal rule for the approximation as it is well known to achieve spectral accuracy for such smooth and periodic integrands. Indeed, *(1) for $\mathbf{x} \notin \mathcal{P}^p$, use of trapezoidal rule to compute two-dimensional "parallel" integrals and a high-order composite Newton-Cotes quadrature for the transverse integral, on a uniform grid $(\mathbf{s}_{ij},t_k) = (ih_{s_1}, jh_{s_2}, kh_t)$ on the parameter space of $\mathcal{P}^p$, yields an overall highly accurate numerical quadrature.*

On the other hand, the case when $\mathbf{x}=\mathbf{x}^p(\mathbf{s},t)$ lies in the integration patch $\mathcal{P}^p$ needs a more careful treatment if high-order accuracy is to be achieved, due to the singular nature of the integrand in Eq. (7). To this end, we begin by using a smooth cut-off function $\eta_{\mathbf{s}}$ satisfying

$$\eta_{\mathbf{s}}(\mathbf{s}') = 1 \text{ for } |\mathbf{s}-\mathbf{s}'| < r_1, \quad \eta_{\mathbf{s}}(\mathbf{s}') = 0 \text{ for } |\mathbf{s}-\mathbf{s}'| \geqslant r_0, \tag{8}$$

to localize the singularity

$$K_p[u](\mathbf{x}) = \int_0^1 \int_{B(\mathbf{s},r_0)} G(\mathbf{x},\mathbf{x}^p(\mathbf{s}',t')) \phi_u^p(\mathbf{s}',t') \eta_{\mathbf{s}}(\mathbf{s}') d\mathbf{s}' dt'$$
$$+ \int_0^1 \int_{[0,1]^2} G(\mathbf{x},\mathbf{x}^p(\mathbf{s}',t')) \phi_u^p(\mathbf{s}',t')(1 - \eta_{\mathbf{s}}(\mathbf{s}')) d\mathbf{s}' dt', \tag{9}$$

where $B(\mathbf{s},r_0)$ denotes the disc of radius r_0 centered at $\mathbf{s}$. The factor $1-\eta$ in the integrand of the second integral in Eq. (9) removes the singularity of the Green's function which can, therefore, be integrated to high-order directly as in the above-described case $\mathbf{x} \notin \mathcal{P}^p$, that is

(2) if $\mathbf{x} \in \mathcal{P}^p$, the integral away from the target point is approximated as in step (1).

To deal with the first integral, on the other hand, we denote the "parallel" integrals corresponding to each t' by $I_u^p(t';\mathbf{x})$, i.e.,
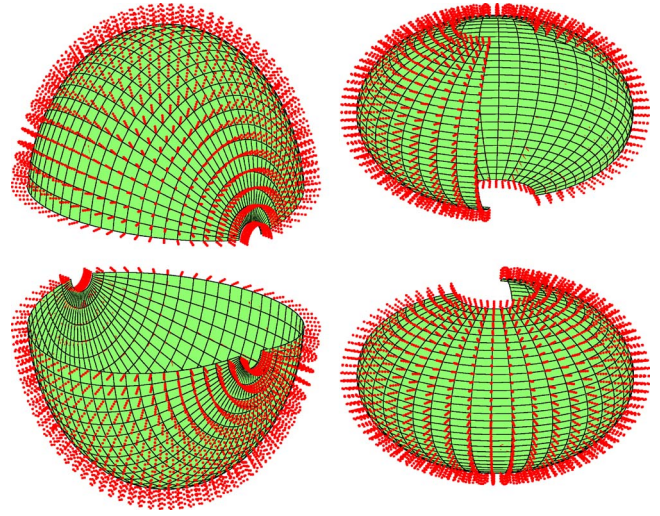
FIG. 2. (Color online) Four patches used to cover the spherical shell and corresponding discretization grids used for the scattering computation.

$$I_u^p(t';\mathbf{x}) = \int_{B(\mathbf{s},r_0)} G(\mathbf{x},\mathbf{x}^p(\mathbf{s}',t')) \phi_u^p(\mathbf{s}',t') \eta_{\mathbf{s}}(\mathbf{s}') d\mathbf{s}' \tag{10}$$

so that it can be rewritten as

$$\int_0^1 I_u^p(t';\mathbf{x}) dt'. \tag{11}$$

As we anticipated, evaluation of $I_u^p(t';\mathbf{x})$ poses difficulties due to the singular character of integrand for $t=t'$ and the "near singularity" when t' is close to t. To overcome these difficulties, the following twofold strategy is used:

(a) A change to polar coordinates (ρ, θ) around the point $\mathbf{s}$ in the $\mathbf{s}'$ plane, transforms the integral in Eq. (10) to

$$I_u^p(t';\mathbf{x}) = \int_0^{2\pi} d\theta \int_0^{r_0} G(\mathbf{x},\mathbf{x}^p(\mathbf{s}+\rho\mathbf{d}_\theta,t'))$$
$$\times \phi_u^p(\mathbf{s}+\rho\mathbf{d}_\theta,t') \eta_{\mathbf{s}}(\mathbf{s}+\rho\mathbf{d}_\theta,t') \rho d\rho,$$

where $\mathbf{d}_\theta = (\cos\theta, \sin\theta)$. The additional factor ρ makes possible the use of a uniform grid in θ (independent of ρ and t') for the approximation of the θ integrals, which can be performed accurately using trapezoidal rule in view of the θ-periodic smooth integrands.

(b) Use of an additional change of variables $\rho=\rho(\tau)$ such that

$$\left. \frac{d^m\rho}{d\tau^m} \right|_{\tau=0} = 0 \quad \text{for } m = 0, \ldots, M \tag{12}$$

[e.g., $\rho(\tau)=\tau^{M+1}$] gives

$$I_u^p(t';\mathbf{x}) = \int_0^{2\pi} d\theta \int_0^{\tau_0} \frac{\rho(\tau)\rho'(\tau)}{|\mathbf{x}-\mathbf{x}^p(\mathbf{s}+\rho(\tau)\mathbf{d}_\theta,t')|} f_u^p(\tau,\theta;t')d\tau, \tag{13}$$

where

FIG. 3. (Color online) Scattering of an incoming plane-wave by a thin spherical shell. (a), (b), and (c) The surface view of the incident, total, and scattered field, respectively; (d), (e), and (f) the corresponding cross-sectional view.

$$f_u^p(\tau, \theta; t')$$

$$= \frac{1}{4\pi} \exp(\iota\kappa |\mathbf{x} - \mathbf{x}^p(\mathbf{s} + \rho(\tau)\mathbf{d}_\theta, t')|)$$

$$\times \phi_u^p(\mathbf{s} + \rho(\tau)\mathbf{d}_\theta, t') \, \eta_\mathbf{s}(\mathbf{s} + \rho(\tau)\mathbf{d}_\theta, t')$$

and $\tau_0 = \rho^{-1}(r_0)$. The integrand in Eq. (13) now has M derivatives in τ that are uniformly bounded and it also vanishes to high order at the end points $\tau = 0$ and $\tau = \tau_0$, so that a high-order integration in τ can be achieved using the trapezoidal rule.

Note that in order to perform the integration in Eq. (13), for each t', we need the density, $\phi_u^p(\mathbf{s} + \rho(\tau)\mathbf{d}_\theta, t')$ on a polar grid $(\tau_{l_1}, \theta_{l_2})$ around $\mathbf{s}$ in the $\mathbf{s}'$ plane (a uniform grid in the θ direction and a ρ-transformed grid in the radial direction). In order to obtain this additional data from the original grid $(\mathbf{s}_{ij}, t_k)$, we use FFT-refined polynomial interpolation procedure described in the following.

(3) Obtain the Fourier coefficients of the density ϕ_u^p for each of the planes $\{(\mathbf{s}, t_k) : \mathbf{s} \in (0,1)^2\}$ using two-dimensional fast Fourier transforms, $\mathcal{FFT}(\phi_u^p)$.

(4) Evaluate the Fourier series on a much finer equis-paced grid $(\mathbf{s}_{ij}/F, t_k)$, where F is the level of refinement (e.g., $F = 2, 4, 8, \ldots$). These evaluations can be obtained efficiently

using two-dimensional fast Fourier transforms, $\mathcal{FFT}^{-1}$. As a result of steps 3 and 4, we achieve a spectrally accurate "trigonometric" interpolation of smooth and periodic density on an F-times finer grid.

(5) Two-dimensional polynomial interpolation of a fixed order R (e.g., bicubic, $R=4$) can be performed locally on the fine grid $(\mathbf{s}_{ij}/F, t_k)$ to obtain the interpolated density values on the polar grid $(\tau_{l_1}, \theta_{l_2})$. The resulting FFT-refined polynomial interpolation scheme is much more accurate for our smooth periodic density than a polynomial interpolation alone [as the error is proportional to $(h_{s_1}/F)^R + (h_{s_2}/F)^R$ rather than simply to $h_{s_1}^R + h_{s_2}^R$]. Finally,

(6) if $\mathbf{x} \in \mathcal{P}^p$ and for integration near the target point, use the interpolated data to evaluate an accurate approximation of the parallel integral (13) in the form of the trapezoidal rule in the transformed variables (τ, θ),

$$I_u^p(t'; \mathbf{x}) \approx \sum_{l_1, l_2} G(\mathbf{x}, \mathbf{x}^p(\mathbf{s} + \rho(\tau_{l_1})\mathbf{d}_{\theta_{l_2}}, t'))$$

$$\times \phi_u^p(\mathbf{s} + \rho(\tau_{l_1})\mathbf{d}_{\theta_{l_2}}, t') \, \eta_\mathbf{s}(\mathbf{s} + \rho(\tau_{l_1})\mathbf{d}_{\theta_{l_2}}, t')$$

$$\times \rho(\tau_{l_1})\rho'(\tau_{l_1})\Delta\tau\Delta\theta.$$

Finally, the integration in the transverse direction

TABLE I. Convergence study. Plane-wave scattering ($\kappa=10$) by a spherical shell (radius=1; thickness=0.025).

Grid size	Unknowns	Samp. density[a]	$\mathcal{E}$	Order
$4\times16\times16\times5$	5 120	1.6	4.6803×10^{-2}	$\cdots$
$4\times32\times32\times5$	20 480	3.2	1.9998×10^{-3}	4.5487
$4\times64\times64\times5$	81 920	6.4	9.2095×10^{-5}	4.4406
$4\times128\times128\times5$	327 680	12.8	1.7774×10^{-6}	5.6953

[a]Number of points per wavelength.

TABLE II. Computational efficiency. Spherical shell geometry.

κ	Thickness	N	β	$\mathcal{E}$	$\mathcal{R}$
2	0.1	5 120	0.325	1.3121×10^{-3}	$\cdots$
4	0.05	20 480	0.419	7.8718×10^{-4}	3.1
8	0.025	81 920	0.490	8.4803×10^{-5}	3.4
16	0.0125	327 680	0.546	4.4295×10^{-4}	3.8

[t'-direction in Eq. (11)] can be handled through a high-order quadrature rule, if carefully partitioned. Indeed the integrand $I_u^p(t';\mathbf{x})$ is precisely a *single layer potential* supported on the surface

$$S_{t'} = \{\mathbf{x}^p(\mathbf{s}',t'):\mathbf{s}' \in (0,1)^2\}$$

and it is therefore smooth for $0\leqslant t'\leqslant t$ and for $t\leqslant t'\leqslant 1$ (in fact, it has a corner type singularity at $t=t'$[27]). Hence we rewrite Eq. (11) as

$$\int_0^t I_u^p(t';\mathbf{x})dt' + \int_t^1 I_u^p(t';\mathbf{x})dt', \tag{14}$$

where each integrand is smooth and can therefore be treated with a high-order composite Newton quadrature with a number Q of points. To this end, however, we need the values of $I_u^p(t';\mathbf{x})$ at points, say $\tilde{t}_k$, other than the original grid points $\{t_k\}$, to evaluate Eq. (14) when t is near 0 or 1 [so that there are enough data points on the grid between 0 and t or t and 1 to enable a high-order quadrature of the integrands in Eq. (14)]. To obtain these extra values, we pursue the following scheme:

(7) *For each $\mathbf{s}_{ij}$ on an equispaced grid, we interpolate the smooth density $\phi_u^p(\mathbf{s}_{ij},t)$ to obtain additional values in t-direction at points $\tilde{t}_k$ near the end points $t=0$ and $t=1$.*

(8) *Perform the parallel local integration in Eq. (13) (again in transformed variables) to obtain the values $I_u^p(\tilde{t}_k;\mathbf{x})$ at these additional points.*

Once we have the additional data, we

(9) *evaluate the full local integral (14) with a high-order Newton-Cotes quadrature in the transverse direction.*

As a result, steps 1–9 provide a high-order quadrature for adjacent interactions. As we said, the use of a partition of unity (6) allows for the integration of the (nonsingular) nonadjacent interactions to be performed to high order with a simple use of the trapezoidal rule. There, however, the central issue is one of computational cost; a strategy to reduce this is described in the following.

B. Nonadjacent interactions: Acceleration

As we noted in Sec. III A, difficulties in integration arise only when the source point is close to the target point [due to the behavior of the kernel $G(\mathbf{x},\mathbf{x}')$]. The above-described integration scheme for adjacent interactions is unnecessary for source points that are sufficiently away from the target point and, in fact, a more efficient strategy can be devised. Indeed, for these, a direct application of the trapezoidal rule yields superalgebraic convergence, on account of the smooth vanishing of the functions $w^p(\mathbf{x})$ in Eq. (6). However such a

procedure results in an $\mathcal{O}(N^2)$ algorithm which may be suitable only for relatively small N. For large N, on the other hand, we propose a variant of the acceleration strategy introduced in Ref. 31, for surface scattering that is suitable to our volumetric problem.

Specifically, as in Ref. 31, our acceleration strategy is based on certain distributions of "equivalent sources" on Cartesian grids. More precisely, we begin by bounding the obstacle by a cube $\mathcal{C}$ of side length A which is then partitioned into L^3 identical cubes $\mathcal{C}_i$ of side $H=A/L$.

We then replace the true sources contained in the cube $\mathcal{C}_i$ by a number M^{equiv} of "equivalent sources" on the faces of $\mathcal{C}_i$ in such a way that the field produced by these equivalent sources coincides, to high-order accuracy, with the field generated by the "true sources" (volume discretization data) in $\mathcal{C}_i$ at all points in space that are "not adjacent" to $\mathcal{C}_i$. Three independent sets $\mathcal{F}_i^l$, $l=1,2,3$ of equivalent sources, comprising both monopoles of intensity $\xi_{i,j}^l$ ($\xi_{i,j}^l G(\mathbf{x},\mathbf{x}_{i,j}^l)$) and dipoles of intensity $\zeta_{i,j}^l$ ($\zeta_{i,j}^l \partial G(\mathbf{x},\mathbf{x}_{i,j}^l)/\partial x_l$), are used for this calculation. These equivalent sources in $\mathcal{F}_i^l$ are placed on an equispaced grid $\mathbf{x}_{i,j}^l$, $j=1,\ldots,M^{\text{equiv}}$ on the union of two faces of $\mathcal{C}_i$ that are parallel to coordinate plane $x_l=0$. The field generated by these equivalent sources

$$K_{na,eq}^{l,\mathcal{C}_i}[u](\mathbf{x}) = \sum_{j=1}^{(1/2)M^{\text{equiv}}} \left(\xi_{i,j}^l G(\mathbf{x},\mathbf{x}_{i,j}^l) \right.$$
$$\left. + \zeta_{i,j}^l \left.\frac{\partial G(\mathbf{x},\mathbf{x}')}{\partial x_l'}\right|_{\mathbf{x}'=\mathbf{x}_{i,j}^l} \right) \tag{15}$$

approximates the field generated by true sources in $\mathcal{C}_i$ within
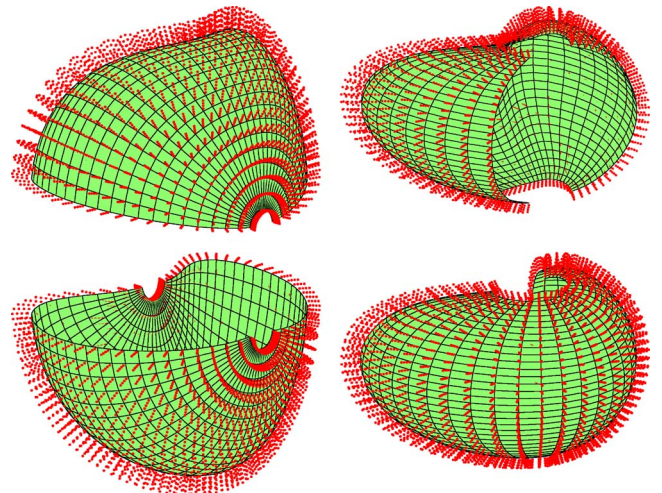
FIG. 4. (Color online) Patches used for the bean shaped scattering geometry and corresponding discretization grids for computation.

A. Anand and F. Reitich: Thin structure acoustic scattering

a prescribed tolerance ϵ, provided M^{equiv} is chosen as[31]

$$M^{\text{equiv}} = \max\left\{\frac{4\kappa^2 A^2}{L^2}, 4c^2(\epsilon)\right\}, \tag{16}$$

where

$$c(\epsilon) = -\frac{\log(\epsilon)}{\log(3)}.$$

Denoting the field generated by the true sources within the cell $\mathcal{C}_i$ by $K^{\mathcal{C}_i}_{\text{true}}[u](\mathbf{x})$, the intensities $\xi^l_{i,j}$ and $\zeta^l_{i,j}$ are then chosen such that the vector formed by the differences $(K^{\mathcal{C}_i}_{\text{true}}[u](\mathbf{x}) - K^{l,\mathcal{C}_i}_{na}[u](\mathbf{x}))$ is minimized in the mean-square norm as $\mathbf{x}$ varies over a number $n^{\text{coll}} \approx 2M^{\text{equiv}}$ collocation points on $\partial \mathcal{S}_i$, where the "supercube" $\mathcal{S}_i$ is union of $\mathcal{C}_i$ and its neighboring cells. Thus (for stability reasons) these intensities are obtained as the least-squares solution of an overdetermined linear system.

Now, recalling the definitions of adjacent and nonadjacent interactions, and replacing $\mathcal{N}_{\mathbf{x}}$ in Eq. (5) with $\mathcal{S}_i$, we have for $\mathbf{x} \in \mathcal{C}_i$,

$$K[u](\mathbf{x}) = K_a[u](\mathbf{x}) + K_{na}[u](\mathbf{x}),$$

where

$$K_a[u](\mathbf{x}) = \int_{\mathcal{S}_i} G(\mathbf{x}, \mathbf{x}')m(\mathbf{x}')u(\mathbf{x}')d\mathbf{x}' \tag{17}$$

and

$$K_{na}[u](\mathbf{x}) = \int_{\dot{\Omega}\backslash\mathcal{S}_i} G(\mathbf{x}, \mathbf{x}')m(\mathbf{x}')u(\mathbf{x}')d\mathbf{x}'. \tag{18}$$

For each $l=1,2,3$, we define

$$K^l_{na,\text{eq}}[u](\mathbf{x}) = \sum_{i=1}^{L^3} K^{l,\mathcal{C}_i}_{na,\text{eq}}[u](\mathbf{x}) - \sum_{\{i|\mathcal{C}_i \subset \mathcal{S}_i\}} K^{l,\mathcal{C}_i}_{na,\text{eq}}[u](\mathbf{x}), \tag{19}$$

which provide accurate approximations to $K_{na}[u]$ throughout the boundary of each cell $\mathcal{C}_i$. The key observation here, of course, is that both summations in Eq. (19), being exact convolutions on a Cartesian grid, can be evaluated efficiently by means of fast Fourier transforms.

Finally, the evaluation of the field values at the true source location from its values at the boundaries of each cell $\mathcal{C}_i$ amounts to solving (Dirichlet) boundary value problems within each $\mathcal{C}_i$. Such Dirichlet problems can be solved uniquely and in a stable manner since the size of the cells can be chosen so that internal resonances do not occur.[31] For efficiency, the field values inside a cell are obtained using a discretized plane wave expansion,[32]

$$K_{na,\text{eq}}[u](\mathbf{x}) = \sum_{k=1}^{n^{\text{wave}}} \psi_k \exp(\iota \kappa \mathbf{d}_k \cdot \mathbf{x}), \tag{20}$$

that approximates $K_{na}[u](\mathbf{x})$. Here $\mathbf{d}_k$ are unit vectors defining directions of wave propagation and $\psi = (\psi_1, \psi_2, \ldots, \psi_{n^{\text{wave}}})$ is a vector of expansion coefficients. Since $K_{na}[u](\mathbf{x})$ does not contain contributions from cells adjacent to $\mathbf{x}$, the expansion (20) converges exponentially

fast with increasing number of wave vectors $\mathbf{d}_k$. The wave direction vectors $\mathbf{d}_k$ are chosen so that they sample the surface of the unit sphere with a sufficient degree of uniformity. The coefficients ψ_k, on the other hand, are chosen such that the expansion (20) matches the field values on the boundary of $\mathcal{C}_i$. The matching procedure thus requires solution of a system of linear equations

$$\mathbf{A}\psi = \mathbf{b}, \tag{21}$$

where the $n^{\text{face}} \times n^{\text{wave}}$-matrix $\mathbf{A}$ consists of elements $\exp(\iota\kappa\mathbf{d}\cdot\mathbf{x})$ evaluated at the equivalent-source points $\mathbf{x} = \mathbf{x}^l_{i,j}$ on the boundary of $\mathcal{C}_i$ and at the discrete direction vectors $\mathbf{d} = \mathbf{d}_k$. The right-hand side of Eq. (21), $\mathbf{b} = (b_1, b_2, \ldots, b_{n^{\text{face}}})$, is a vector of known values of $K_{na}[u]$ at all the equivalent-source points $\mathbf{x}^l_{i,j}$ on the boundary of $\mathcal{C}_i$. Again here, due to stability considerations, the problem is set up such that the associated linear system (21) is overdetermined.

Note that the identical geometry of cells $\mathcal{C}_i$ enables us to compute the QR-factorization of $\mathbf{A}$ only once and to store it for repeated use.

IV. ACCURACY AND EFFICIENCY

The above-presented integration schemes for computing adjacent and nonadjacent interactions are designed with the dual goal of achieving high order accuracy and computational efficiency. In this section we provide a brief summary of the arguments that substantiate these claims; numerical examples that further confirm these derivations are presented in Sec. V. Specifically, first, in Sec. IV A, we show that our numerical solution indeed converges to the exact solution with a tunable convergence order. In Sec. IV B then, we provide a derivation that delivers an optimal choice (to minimize computational effort) for the number L^3 of cells to be used in the acceleration procedure (cf. Sec. III B).

A. Accuracy: Error analysis

A careful derivation of an estimate for the error incurred by the algorithm presented in Sec. III can be performed in a rather straightforward manner, much as in other standard analyses of schemes for the solution of integral equations (see, e.g., Ref. 33). The basic estimate relates to that of the quadrature error, as an appeal to the uniform invertibility of the discretized operators then translates this into the actual error in the computed solution.

In our case, the quadrature error can be easily estimated, through bounds on the error incurred at each step of our procedure. Briefly, for instance, the error in the trapezoidal rule of steps 1 and 2 presented earlier is proportional to h_s^k where k denotes the degree of smoothness of the field *within* Ω, and h_s denotes the parallel grid spacing. Similarly, the error in the composite Newton-Cotes quadrature is no larger than a constant multiple of h_t^Q, where h_t is the grid size in the transverse direction. The error in the Fourier/polynomial interpolation of steps 3–5, on the other hand, is bounded by a constant times $(h_s/F)^R$ and, similarly, that in step 7 by h_t^R. The adjacent, parallel integration of steps 6 and 8, in turn, incur an error proportional to h_s^k and, finally, the transverse

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

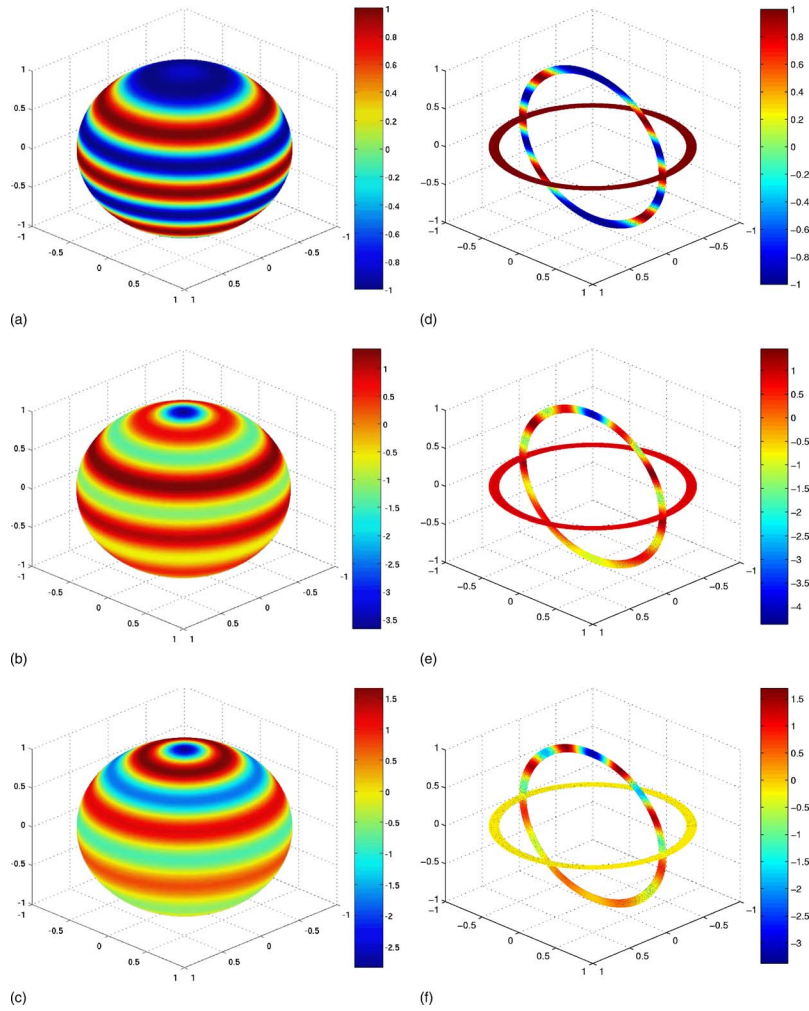A. Anand and F. Reitich: Thin structure acoustic scattering   2509
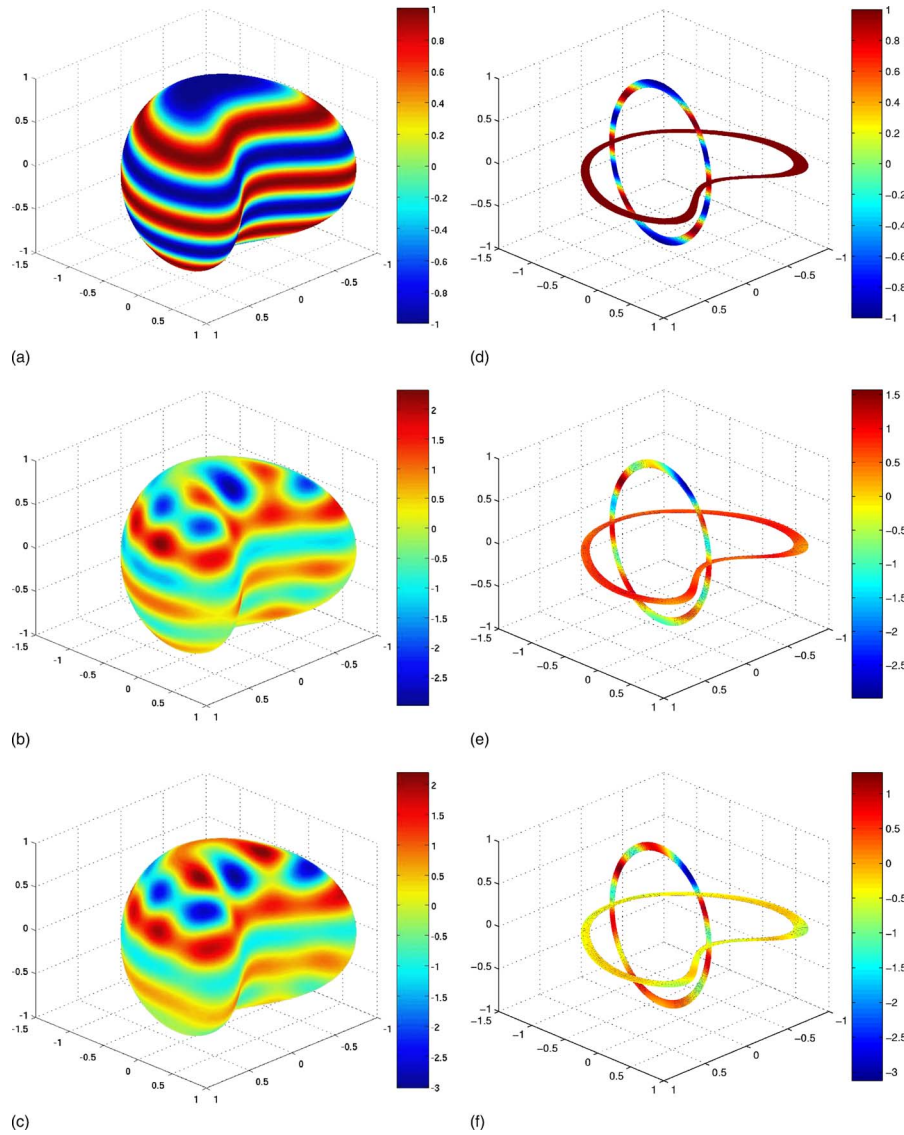
FIG. 5. (Color online) Scattering of an incoming plane wave by a bean-shaped thin scatterer. (a), (b), and (c) The surface view of the incident, total, and scattered field, respectively; (d), (e), and (f) the corresponding cross-sectional view.

integration in step 9 approximates the true value within $O(h_t^Q)$. Combining these observations, and letting F be such that $h_s/F < h_t$, we obtain

$$\text{Max Error} \leqslant \text{Const}(h_s^k + h_t^{\min(Q,R)}).$$

The above-presented analysis of the numerical scheme establishes the high-order convergence of the algorithm. The remaining issue of computational efficiency is analyzed in the following.

B. Efficiency: Operation count

The number of floating point operations required to compute $K[u]$ depends on L and M^{equiv} (cf. Sec. III B), in addition to the number N of volume discretization points. The latter must, of course, be chosen so as to correctly sample the field, on the scale of the wavelength of radiation, as well as to resolve material/geometrical variations. The number M^{equiv}, in turn, should guarantee that the equivalent

field is also well-resolved. Since M^{equiv} is the number of points on a face of a cell C_i, of area $(A/L)^2$, we must have

$$M^{\text{equiv}} \sim \kappa^2 A^2/L^2. \tag{22}$$

The number L, on the other hand, is at our disposal and can therefore be chosen so to minimize the computational effort to attain a prescribed accuracy ϵ. To this end, we must evaluate the cost of each step of our procedure, as a function of L; the optimal choice then can be easily (and, explicitly) derived as that which minimizes the resulting overall cost.

To evaluate the cost of each step, we first note that, for a thin structure, only $\mathcal{O}(L/(\kappa A) \times L^2) = \mathcal{O}(L^3/(\kappa A))$ out of a total of L^3 cells C_i contain source points; the rest are empty. As a result,

$$\mathcal{O}(\kappa A N/L^3) \tag{23}$$

discretization points lie, on average, in each of the nonempty cells.
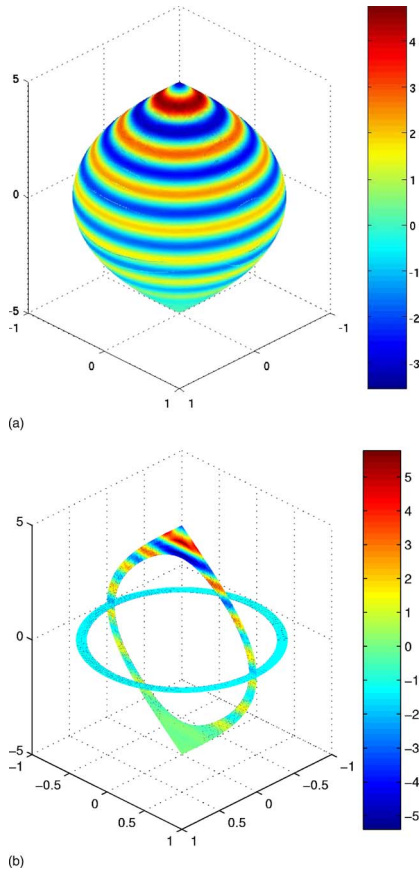
FIG. 6. (Color online) Scattering of an incoming plane wave by a thin scatterer with singularities. (a) and (b) The surface view and cross-sectional view of the computed scattered field, respectively, corresponding to an incident plane wave $\exp(\iota\kappa x_3)$ with $\kappa=5$.

Thus, the total number of operations required for the local integration (cf. Sec. III A) is as follows.

(1) Cost of adjacent integrations $=\mathcal{O}(N \times \kappa A N/L^3)$ $=\mathcal{O}(\kappa A N^2/L^3)$.

The interpolations necessary to allow for this integration (steps 3, 4, 5, and 7 in Sec. III A), on the other hand, cost no more than

(2) Cost of interpolation $=\mathcal{O}(N \log N)$ on account of the FFT refinement.

The cost of the acceleration scheme, finally, is largely determined by the need to perform three convolutions through FFTs on three-dimensional grids of size $L^3 M^{\text{equiv}}$ (L planes containing $L^2 M^{\text{equiv}}$ equivalent sources each).

(3) Cost of evaluation of equivalent source interactions $=\mathcal{O}(L^3 M^{\text{equiv}} \log(L^3 M^{\text{equiv}}))$.

In addition to this cost, we must also consider that of the corrections in Eq. (19) corresponding to cells $\mathcal{C}_i \subset \mathcal{S}_i$, and that of the least-squares problems that determine the equivalent sources and the coefficients of plane wave expansions [cf. Eq. (21)]. The former can be computed with three-dimensional FFTs of size M^{equiv} on each cell, for a total cost of

(4) Cost of corrections to equiv. source interactions $=\mathcal{O}((L^3/(\kappa A)) M^{\text{equiv}} \log(M^{\text{equiv}}))$, while the latter demands the QR factorization of a single matrix of size M^{equiv} and the evaluation of the right-hand sides, for a cost of

(5) Cost of solution of (all) least-squares problems $=\mathcal{O}((M^{\text{equiv}})^3)+\mathcal{O}(NM^{\text{equiv}})$.

This last equality uses the fact that M^{equiv} is no bigger that the number of points per non-empty box [cf. Eq. (23)], which follows from Eq. (22) and the fact that N must be at least as large as the acoustic volume $\kappa^2 A^2$ of the scatterer,

$$
M^{\text{equiv}} \sim \mathcal{O}\left(\frac{\kappa A}{L}\frac{\kappa A}{L}\right) \leq \mathcal{O}\left(\frac{\kappa A}{L}\left(\frac{\kappa A}{L}\right)^2\right)
$$

$$
\leq \mathcal{O}\left(\frac{\kappa A}{L}\frac{N}{L^2}\right)=\mathcal{O}\left(\kappa A \frac{N}{L^3}\right). \tag{24}
$$

Collecting these costs leads to an expression that, for each given N, can be minimized in L. A simple calculation then reveals that the optimal value of L is

$$
L_{\text{opt}}=\begin{cases} \mathcal{O}(N^{(4-\beta)/8}) & \text{if } 4/5 \leq \beta \leq 1 \\ \mathcal{O}(N^{(4+\beta)/12}) & \text{if } \beta < 4/5, \end{cases} \tag{25}
$$

where

$$
\beta(N, \kappa A) = \log(\kappa^2 A^2)/\log(N) \tag{26}
$$

is a relative measure of the refinement of the grid. With these choices we have

$$
\text{Optimal Cost}=\begin{cases} \mathcal{O}(N^{(7\beta+4)/8} \log N) & \text{if } 4/5 \leq \beta \leq 1 \\ \mathcal{O}(N^{(\beta+4)/4} \log N) & \text{if } \beta < 4/5. \end{cases}
$$

(27)

Note that in the first case, corresponding to coarser discretizations, the operation count lies between $\mathcal{O}(N^{6/5} \log N)$ and $\mathcal{O}(N^{11/8} \log N)$. In the more relevant case of $\beta < 4/5$, i.e., for more refined grids, the operation is lower and it approaches $N \log N$ with increasing refinement. In any case, as we said, the count is never higher than $\mathcal{O}(N^{11/8} \log N)$.

V. NUMERICAL RESULTS

In this section, we present numerical experiments for various thin structures in three dimensions. In all of the examples, the linear systems resulting from the discretization of Eq. (2) are solved by means of the iterative solver GMRES[34] in fully complex arithmetic.

The first example involves scattering computations for a thin spherical shell of refractive index $\sqrt{2}$ resulting from the incoming wave

$$
u^i(\mathbf{x}) = e^{\iota\kappa\mathbf{d}\cdot\mathbf{x}} \tag{28}
$$

for which an exact solution can be computed. Four patches (see Fig. 2) were used to describe the thin spherical shell via smooth local parametrizations and the parameter values $F=4$, $Q=5$, and $R=4$ were used for the numerical computations. Recall that the parameter F is the level of refinement used in the trigonometric interpolation (steps 3 and 4), Q is the number of points used for the Newton-Cotes quadrature (steps 1, 2, and 9), and $R1$ is the degree of the interpolating polynomials (steps 5 and 7). The surface and cross-sectional views of the computed fields in the case $\mathbf{d}=(0,0,1)$ are presented in Fig. 3.

A. Anand and F. Reitich: Thin structure acoustic scattering 2511
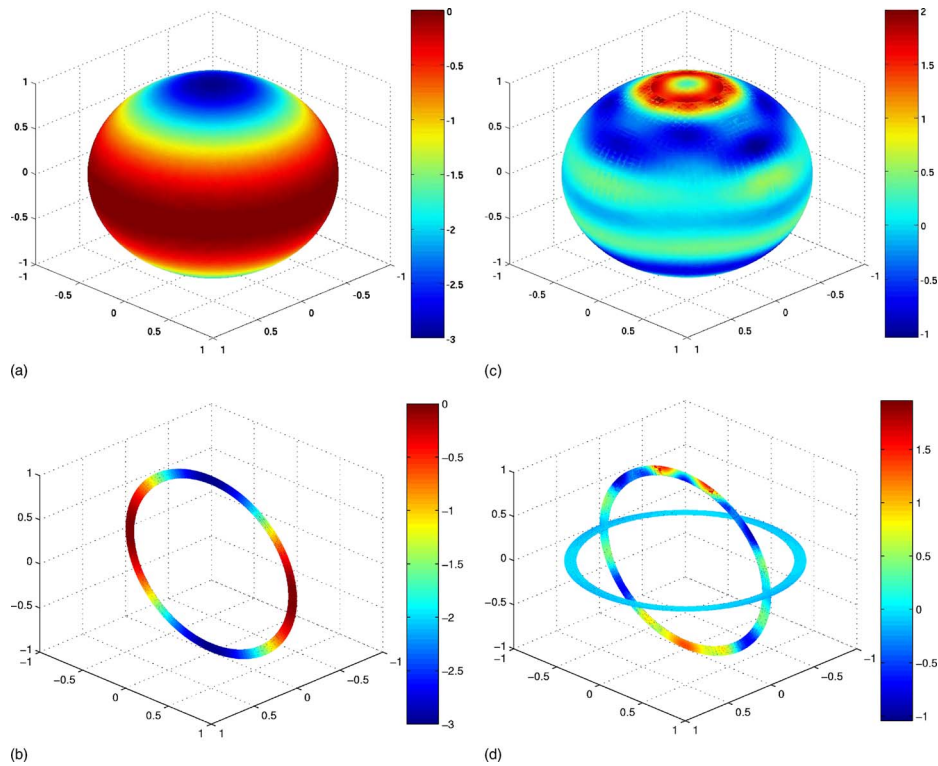
FIG. 7. (Color online) Variable index example. (a) and (b) The surface and cross-sectional views of the refractive index. (c) and (d) The surface view and cross-sectional view of the computed scattered field, respectively, corresponding to an incident plane wave $\exp(\iota \kappa x_3)$ with $\kappa = 10$.

Numerical results on the maximum relative (near-field) error $\mathcal{E}$,

$$\mathcal{E} = \frac{\max_\Omega |u^{\text{exact}} - u^{\text{computed}}|}{\max_\Omega |u^{\text{exact}}|},$$

for the example of Fig. 3 are presented in Table I. Clearly, these results confirm the high-order convergence of the scheme (cf. Sec. IV A). In Table II, on the other hand, we present a numerical study of the computational complexity associated with this example. To this end we consider a sequence of scattering problems with increasing frequencies and wherein the thickness of the domain is correspondingly reduced so as to continue to satisfy the thinness assumption (4). Table II shows that the computational effort grows as predicted by Eq. (27). To display this, the growth in the computational cost for successive doublings of the wave number, and consequent quadrupling of the number of points, is measured by

$$\mathcal{R} = \frac{(T_{2\kappa}/\log 4N)^{1/q_{2\kappa}}}{(T_\kappa/\log N)^{1/q_\kappa}}.$$

Here T_κ and $T_{2\kappa}$ are the computational times corresponding to κ and 2κ, respectively, and $q_\kappa = (4 + \beta_\kappa)/4$ and $q_{2\kappa} = (4 + \beta_{2\kappa})/4$ denote the powers in the count (27). As predicted the ratios $\mathcal{R}$ do not exceed 4, the ratio of the number of points.

Our algorithm, of course, is not constrained by the geometry of the scattering object. In fact, the only requirement to handle any given geometry is that suitable local parametrizations be constructed. To demonstrate this, we present two additional examples involving a bean-shaped thin scatterer and a thin volume geometry with conic singularities, with $n^2 = 2$. The former geometry corresponds to the parametrization

TABLE III. (a) Maximum relative near field error (as compared with the analytical solution) for the scattering configuration in Fig. 8 with N discretization points. (b) Computational time per iteration for incident plane waves of increasing frequency (and corresponding increases in the number of unknowns).

(a)				(b)			
						Time/iter.	
N	Samp. density[a]	Max error	Order	κa	N	(s)	Order
3×16^2	4	$2.3795E{-}1$		4	3×32^2	7	
3×32^2	8	$1.8715E{-}3$	6.99	8	3×64^2	35	1.2
3×64^2	16	$2.6252E{-}5$	6.16	16	3×128^2	151	1.1
3×128^2	32	$1.3674E{-}7$	7.58	256	3×256^2	688	1.1

[a]Number of points per wavelength

A. Anand and F. Reitich: Thin structure acoustic scattering

FIG. 8. (Color online) (a) The decomposition of an inhomogeneity (a circular cylinder with radius a and constant refractive index $n=2$) into two thin-volume patches and one bulk patch, with the associated discretization mesh. (b) The near field intensity of the solution u due to an incident plane wave traveling from lower-left to upper-right with $\kappa a = 4$.

$$\frac{x_1^2}{a^2} + \frac{\left(\alpha_1 A \cos\dfrac{\pi x_1}{A} + x_2\right)^2}{b^2\left(1 - \alpha_2 \cos\dfrac{\pi x_1}{A}\right)} + \frac{x_3^2}{c^2\left(1 - \alpha_3 \cos\dfrac{\pi x_1}{A}\right)}$$

$$= A^2(1 - Bt)^2$$

with $a=1.0$, $b=1.0$, $c=1.0$, $\alpha_1=0.3$, $\alpha_2=0.4$, $\alpha_3=0.1$, $A=1.0$, $B=0.1$, and $0 \leq t \leq 1$ (see Fig. 4). The field computations for this configuration [$\kappa=10$, $\mathbf{d}=(0,0,1)$] are presented in Fig. 5. Results for the singular geometry

$$x_1^2 + x_2^2 + \sin^2\left(\frac{\pi x_3}{2c}\right) = (1 - Bt)^2$$

with $c=5.0$, $B=0.1$, and $0 \leq t \leq 1$, in turn, are presented in Fig. 6 [$\kappa=5$, $\mathbf{d}=(0,0,1)$].

By design, our algorithm is also not restricted to consideration of constant or piecewise constant refractive indices. An example with variable refractive index is included in Fig. 7 for a spherical shell scatterer. The refractive index is given by

$$n(r, \phi, \theta) = 1 + \cos^2 \theta$$

and the incidence is a plane wave with $\kappa=10$ and $\mathbf{d}=(0,0,1)$.

VI. DISCUSSION

Our present work concentrates on the specialized treatment of regions near material discontinuities, which we view as thin scatterers. Here, however, we briefly discuss the application of these ideas to the general case, wherein we couple our new schemes to the more classical approaches, such as those described in Ref. 26, to deliver fast and accurate solutions in the presence of discontinuities. The basic premise here is that, as we mentioned, these latter spectral methodologies deliver highly accurate solutions whenever the scattering media vary smoothly. In Fig. 8 we display some preliminary results of the integration of the spectral scheme proposed in Ref. 26 with our new thin-volume scattering solver. Figure 8(a) shows the "partition of unity" used here to smoothly separate the "bulk" of the domain, to be treated with the method of Ref. 26, from the thin "boundary volumes" to which the schemes presented in this text are applied. The interaction between the bulk and boundary volumes are effected via FFTs and equivalent source calculations (see Sec. III B). Table III, on the other hand, shows that high-order accuracy is indeed attained, and that the cost grows only linearly with problem size. A thorough analysis and further applications of such integrated algorithms, however, are left for future work.

The numerical experiments indicate that the number of iterations required by GMRES to solve the linear system arising from the discretized integral equation, without any preconditioning, increases quadratically with the problem size (e.g., κa). For example, 23 iterations were needed in the above-presented example for $\kappa a = 4$. Although a considerable research effort has been devoted to the development of effective preconditioning techniques for such scattering problems,[35,36] it still cannot be considered completely solved and remains the subject matter of much ongoing research.

VII. CONCLUSIONS

We have presented a numerical scheme for scalar thin-structure scattering simulations that can deliver highly accurate results with a reduced operation count. The method is based on an accurate numerical representation of the singular operators that arise in the integral equation formulation and on FFT-accelerated evaluations. The former are attained through specialized high-order quadratures, while the latter rely on the evaluation of equivalent sources on Cartesian grids.

The theoretical and numerical results presented here demonstrate the viability of the approach in the context of acoustic scattering off thin structures that may be of independent interest (e.g., material coatings). As follows from our discussion, however, the basic ideas behind our schemes can be readily extended to other wave models (electromagnetics, elasto-dynamics). More importantly perhaps, we have also explained how our new algorithms can provide an effective means to upgrade existing FFT-based numerical methods to allow these to accurately handle material discontinuities, by smoothly separating the scatterer into an interior bulk and thin boundary volumes. Indeed, as we said, any method that can efficiently and accurately compute the scatter off

smoothly varying media can be used to treat the bulk, and interfaced with the thin volume solver presented here to compute the fields in the vicinity of discontinuities. As we have shown, this interface can be made seamless and in a manner that provides high-order convergence without compromising efficiency.

ACKNOWLEDGMENTS

[1] G. Cohen and P. Joly, "Construction and analysis of fourth order finite difference schemes for the acoustic wave equations in nonhomogeneous media," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **33**, 1266–1302 (1996).

[2] A. Bamberger, R. Glowinski, and Q. H. Tran, "A domain decomposition method for the acoustic wave equation with discontinuous coefficients and grid change," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **34**, 603–639 (1997).

[3] B. Gustafsson and P. Wahlund, "Time compact difference methods for wave propagation in discontinuous media," SIAM J. Sci. Comput. (USA) **26**, 272–293 (2004).

[4] A. Kirsch and P. Monk, "Convergence analysis of a coupled finite-element and spectral method in acoustic scattering," IMA J. Numer. Anal. **10**, 425–447 (1990).

[5] W. Rachowicz and L. Demkowicz, "An hp-adaptive finite element method for electromagnetics. 1. Data structure and constrained approximation," Comput. Methods Appl. Mech. Eng. **187**, 307–335 (2000).

[6] A. Kirsch and P. Monk, "An analysis of the coupling of finite-element and nystrom methods in acoustic scattering," IMA J. Numer. Anal. **14**, 523–544 (1994).

[7] Q. H. Liu, "Large scale simulations of electromagnetic and acoustic measurements using the pseudospectral time-domain method (pstd) algorithm," IEEE Trans. Geosci. Remote Sens. **37**, 917–226 (1999).

[8] B. Yang, D. Gottlieb, and J. S. Hesthaven, "Spectral simulations of electromagnetic wave scattering," J. Comput. Phys. **134**, 216–230 (1997).

[9] Q. H. Liu, "The pseudospectral time-domain (pstd) algorithm for acoustic waves in absorptive media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1044–1055 (1998).

[10] T. D. Mast, L. P. Souriau, D. L. D. Liu, M. Tabei, A. I. Nachman, and R. C. Waag, "A k-space method for large scale-models of wave propagation in tissue," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 341–354 (2001).

[11] M. Tabei, T. D. Mast, and R. C. Waag, "A k-space method for coupled first-order acoustic propagation equations," J. Acoust. Soc. Am. **111**, 53–63 (2002).

[12] N. N. Bojarski, "The k-space formulation of the scattering problem in the time domain," J. Acoust. Soc. Am. **72**, 570–884 (1982).

[13] N. N. Bojarski, "The k-space formulation of the scattering problem in the time domain: An improved single propagator formulation," J. Acoust. Soc. Am. **77**, 826–831 (1985).

[14] J. Lu, J. Pan, and B. Xu, "Time-domain calculation of acoustical wave propagation in discontinuous media using acoustical wave propagator with mapped pseudospectral method," J. Acoust. Soc. Am. **118**, 3408–3419 (2005).

[15] A. Dutt and V. Rokhlin, "Fast Fourier transform for nonequispaced data," SIAM J. Sci. Comput. (USA) **14**, 1368–1393 (1993).

[16] N. Nguyen and Q. H. Liu, "The regular Fourier matrices and non-uniform fast Fourier transforms," SIAM J. Sci. Comput. (USA) **21**, 283–293 (1999).

[17] Q. H. Liu, X. M. Xu, B. Tian, and Z. Q. Zhang, "Applications of non-uniform fast Fourier transform algorithms in numerical solutions of differential and integral equations," IEEE Trans. Geosci. Remote Sens. **38**, 1551–1560 (2000).

[18] D. Gottlieb, C. W. Shu, A. Solomonoff, and H. Vandeven, "On the Gibbs phenomenon. i. Recovering exponential accuracy from the Fourier partial sum of a nonperiodic analytic function," J. Comput. Appl. Math. **43**, 81–98 (1992).

[19] S. A. Sarra, "Spectral methods with postprocessing for numerical hyperbolic heat transfer," Numer. Heat Transfer, Part A **43**, 717–730 (2003).

[20] J. P. Boyd, "Trouble with Gegenbauer reconstruction for defeating Gibbs' phenomenon: Runge phenomenon in the diagonal limit of Gegenbauer polynomial approximations," J. Comput. Phys. **204**, 253–264 (2005).

[21] X. M. Xu and Q. H. Liu, "Fast spectral method for acoustic scattering problems," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 522–529 (2001).

[22] P. Zwamborn and P. D. van den Berg, "The three-dimensional weak form of the conjugate gradient fft method for solving scattering problems," IEEE Trans. Microwave Theory Tech. **40**, 1757–1766 (1992).

[23] G. X. Fan and Q. H. Liu, "The CGFFT method with a discontinuous FFT algorithm," Microwave Opt. Technol. Lett. **29**, 47–49 (2001).

[24] G. Liu and S. D. Gedney, "High-order nystrom solution of the volume—efie for te-wave scattering," Electromagnetics **21**, 1–13 (2001).

[25] A. Zhu and S. D. Gedney, "A quadrature-sampled pre-corrected FFT method for the electromagnetic scattering from inhomogeneous objects," IEEE Antennas Wireless Propag. Lett. **2**, 50–53 (2003).

[26] E. M. Hyde and O. P. Bruno, "A fast, higher-order solver for scattering by penetrable bodies in three dimensions," J. Comput. Phys. **202**, 236–261 (2005).

[27] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, 2nd ed. (Springer, Berlin, 1998).

[28] P. Werner, "Zur mathematischen theorie akustischer wellenfelder (On the mathematical theory of acoustic wave fields)," Arch. Ration. Mech. Anal. **6**, 231–260 (1960).

[29] P. Werner, "Beugungsprobleme der mathematischen akustik (Diffraction problems of mathematical acoustics)," Arch. Ration. Mech. Anal. **12**, 155–184 (1963).

[30] P. A. Martin, "Acoustic scattering by inhomogeneous obstacles," SIAM J. Appl. Math. **64**, 297–308 (2003).

[31] O. P. Bruno and L. A. Kunyansky, "A fast, high-order algorithm for the solution of surface scattering problems: Basic implementation, tests and applications," J. Comput. Phys. **169**, 80–110 (2001).

[32] L. B. Felsen and N. Markuvitz, *Radiation and Scattering of Waves* (Prentice-Hall, New York, 1973).

[33] R. Kress, *Linear Integral Equations* (Springer, Berlin, 1989).

[34] Y. Saad and M. H. Schultz, "Gmres: A generalized minimal residual algorithm for solving non-symmetric linear systems," SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **7**, 856–869 (1986).

[35] J. Lee, J. Zhang, and C. C. Lu, "Incomplete Lu preconditioning for large scale dense complex linear systems from electromagnetic wave scattering problems," J. Comput. Phys. **185**, 158–175 (2003).

[36] E. M. Hyde and O. P. Bruno, "An efficient, preconditioned, high-order solver for scattering by two-dimensional inhomogeneous media," J. Comput. Phys. **200**, 670–694 (2004).

Generation of shock-free pressure waves in shaped resonators by boundary driving

C. Luo, X. Y. Huang,[a) and N. T. Nguyen
School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore

Investigation of high amplitude pressure oscillations generated by boundary driving in shaped resonators has been carried out both theoretically and experimentally. In the theoretical modeling, the acoustic resonance in an axisymmetric resonator is studied by the Galerkin method. The resonator is exponentially expanded and the boundary driving is provided by a piston at one end. The pressure wave forms, amplitudes, resonance frequencies, and ratio of pressures at the two ends of the resonator are calculated for various expansion flare constants and driving strengths. These results are partially compared with those generated by shaking the resonator. They are also verified in the experiment, in which an exponentially expanded resonator is connected to a speaker box functioning as the piston. The experiment is further extended to a horn-shaped resonator with a rectangular cross section. The boundary driving in this case is generated by a circular piezoelectric disk, which forms one sidewall of the resonator cavity. The characteristics of axisymmetric resonators, such as the resonance frequency and amplitude ratio of pressures at the two ends, are observed in this low aspect ratio rectangular resonator with the sidewall driving. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2713716]

I. INTRODUCTION

High amplitude pressures can be generated by shaking shaped resonators whose cross-section areas are expanded along the axial dimensions. This has been demonstrated by both experiments[1] and theoretical modeling.[2–6] In the experiments conducted by Lawrenson *et al.*,[1] standing wave overpressures in excess of 340% ambient pressure were recorded by shaking shaped resonators. Ilinskii *et al.*[2] developed a one-dimensional model to analyze the nonlinear standing waves in shaped resonators. The large amplitude pressures, wave form distortion, and resonance frequency shift were calculated and compared with the experimental results. This one-dimensional model was later improved by including the shear viscosity term in the momentum equation.[3,4] Ilinskii *et al.*[3] studied the energy losses and the quality factors of shaped resonators. Luo *et al.*[4] investigated the effect of resonator dimensions. An analytical study was conducted by Hamilton *et al.*[5] to predict and explain the dependence of the nonlinear frequency response on resonator shapes. Erickson and Zinn[6] proposed a procedure to solve the one-dimensional wave equation for exponentially expanded resonators by the Galerkin method. The resonance characteristics are discussed in terms of expansion flare constant. Other than shaking the shaped resonators, generation of high amplitude pressure waves in a tube with an array of Helmholz resonators are also reported.[7,8]

The driving method to generate high amplitude pressures in shaped resonators has been so far focused on shaking the entire resonators, probably because the entire shaking is more efficient in power delivery[1] compared to other meth-

ods, such as piston driving. In order to broaden the potential of this technology in other engineering applications, such as miniature or microfluidic devices,[9,10] it is necessary to explore other driving mechanisms for various resonator configurations. The boundary driving in these applications may be a better way than shaking to generate useful dynamic pressures, which motivated the present study. In this study, generation of high amplitude pressure in shaped resonators by boundary driving was investigated both theoretically and experimentally. The theoretical modeling using Galerkin's method was performed on an axisymmetric resonator with one end connected to a piston. Calculated results, including the resonance frequencies, pressure amplitudes at two ends of the resonator, and their ratios were compared with measured data from the experiment, which was conducted on an axial symmetric resonator. The experimental study was further extended to a horn-shaped resonator with low aspect ratio rectangular cross sections. The boundary driving in this case was from a piezoelectric disk which formed the sidewall of the resonator cavity. The resonance frequency and pressures were measured.

II. MODELING

A. Basic equations and boundary conditions

We consider an axisysmmetric resonator shown in Fig. 1. The resonator has length *l* and radius $r(x)$ which is a function of *x*. The cross-sectional area of the resonators is expanded exponentially along its axis. The resonator is sealed at the small end ($x=0$) and connected to a piston at the other end ($x=l$). The configuration of the resonator is very similar to the one previously used by Luo *et al.*,[4] except that there is no body shaking force in the present case. The
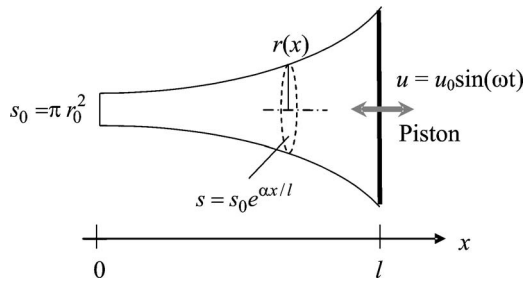
FIG. 1. The exponentially expanded resonator with a piston driving at one end.

dimensionless nonlinear wave equation previously used by Luo et al.[4] for the resonator under shaking is written here:

$$
\frac{\partial^2 \Phi}{\partial T^2} - \frac{1}{\Omega^2 \pi^2} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi}{\partial X} - \frac{1}{\Omega^2 \pi^2} \frac{\partial^2 \Phi}{\partial X^2} + \frac{2}{\Omega} \frac{\partial^2 \Phi}{\partial X \partial T} \frac{\partial \Phi}{\partial X}
$$
$$
+ \frac{(\gamma-1)}{\Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi}{\partial T} \frac{\partial \Phi}{\partial X} + \frac{(\gamma-1)}{\Omega} \frac{\partial \Phi}{\partial T} \frac{\partial^2 \Phi}{\partial X^2}
$$
$$
- \frac{G_B}{\pi^3 \Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial^2 \Phi}{\partial T \partial X} - \frac{G_B}{\pi^3 \Omega} \frac{\partial^3 \Phi}{\partial T \partial X^2} + \frac{G_S}{R\Omega^{1/2}} \frac{\partial \Phi}{\partial T}
$$
$$
= F_{\text{shaking}}, \tag{1}
$$

where

$$
F_{\text{shaking}} = -\frac{1}{\Omega^2} \left\{ A \frac{\partial \Phi}{\partial X} + \Omega \frac{\partial A}{\partial T} X + (\gamma-1) \left\{ \frac{1}{R^2} \frac{dR^2}{dX} \right\} \right.
$$
$$
\left. \times \frac{\partial \Phi}{\partial X} AX + (\gamma-1) \frac{\partial^2 \Phi}{\partial X^2} AX \right\} \tag{2}
$$

consists of the terms related to the dimensionless shaking acceleration A. In Eqs. (1) and (2),

$$
X = \frac{x}{l}, \quad T = \omega t, \quad S = \frac{\pi r^2}{l^2}, \quad R = \frac{r}{l},
$$
$$
\omega_0 = \frac{\pi c_0}{l}, \quad A = \frac{a}{l\omega_0^2}, \tag{3}
$$

$$
G_B = \frac{(\zeta + 4\eta/3)\pi\omega_0}{\rho_0 c_0^2}, \quad G_S = \sqrt{\frac{2\eta}{\pi\rho_0 c_0 l}},
$$
$$
\Omega = \frac{\omega}{\omega_0}, \quad \Phi = \frac{\varphi}{l^2 \omega_0}, \tag{4}
$$

where c_0 is the speed of sound, ζ and η are coefficients of viscosity, $a = a_0 \sin(\omega t)$ is acceleration associated with the shaking, and φ is velocity potential. By setting $A = 0$, F_{shaking} will be zero in Eq. (1), and the dimensionless wave equation for the present case is obtained as

$$
\frac{\partial^2 \Phi}{\partial T^2} - \frac{1}{\Omega^2 \pi^2} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi}{\partial X} - \frac{1}{\Omega^2 \pi^2} \frac{\partial^2 \Phi}{\partial X^2} + \frac{2}{\Omega} \frac{\partial^2 \Phi}{\partial X \partial T} \frac{\partial \Phi}{\partial X}
$$
$$
+ \frac{(\gamma-1)}{\Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi}{\partial T} \frac{\partial \Phi}{\partial X} + \frac{(\gamma-1)}{\Omega} \frac{\partial \Phi}{\partial T} \frac{\partial^2 \Phi}{\partial X^2}
$$
$$
- \frac{G_B}{\pi^3 \Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial^2 \Phi}{\partial T \partial X} - \frac{G_B}{\pi^3 \Omega} \frac{\partial^3 \Phi}{\partial T \partial X^2}
$$
$$
+ \frac{G_S}{R\Omega^{1/2}} \frac{\partial \Phi}{\partial T} = 0. \tag{5}
$$

The piston velocity at $x = l$ in Fig. 1 is $u(t) = u_0 \sin(\omega t)$, and its dimensionless form is

$$
U(T) = U^* \sin(T), \tag{6}
$$

where $U^* = u_0 / l\omega_0$ is the velocity amplitude of the piston, indicating the driving strength. The boundary conditions in dimensionless forms are therefore

$$
\left. \frac{\partial \Phi}{\partial X} \right|_{X=0} = 0, \quad \left. \frac{\partial \Phi}{\partial X} \right|_{X=1} = U^* \sin(T). \tag{7}
$$

Since the piston is oscillating at the right end, the resonator length l in the present case is measured from $x = 0$ to the average position of piston. The nonlinear wave equation (5), together with boundary conditions (7) may be solved by assuming that Φ consists of two parts, i.e.,

$$
\Phi = \Phi_A + \Phi_B, \tag{8}
$$

where Φ_A is a new unknown function to be solved, which satisfies homogeneous boundary conditions

$$
\left. \frac{\partial \Phi_A}{\partial X} \right|_{X=0} = 0, \quad \left. \frac{\partial \Phi_A}{\partial X} \right|_{X=1} = 0, \tag{9}
$$

and Φ_B can be chosen to satisfy the nonhomogeneous boundary conditions (7). By substituting Eq. (8) into Eq. (5), one obtains

$$
\frac{\partial^2 \Phi_A}{\partial T^2} - \frac{1}{\Omega^2 \pi^2} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi_A}{\partial X} - \frac{1}{\Omega^2 \pi^2} \frac{\partial^2 \Phi_A}{\partial X^2} + \frac{2}{\Omega} \frac{\partial^2 \Phi_A}{\partial X \partial T} \frac{\partial \Phi_A}{\partial X}
$$
$$
+ \frac{(\gamma-1)}{\Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi_A}{\partial T} \frac{\partial \Phi_A}{\partial X} + \frac{(\gamma-1)}{\Omega} \frac{\partial \Phi_A}{\partial T} \frac{\partial^2 \Phi_A}{\partial X^2}
$$
$$
- \frac{G_B}{\pi^3 \Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial^2 \Phi_A}{\partial T \partial X} - \frac{G_B}{\pi^3 \Omega} \frac{\partial^3 \Phi_A}{\partial T \partial X^2}
$$
$$
+ \frac{G_S}{R\Omega^{1/2}} \frac{\partial \Phi_A}{\partial T} = F_{\text{piston}}, \tag{10}
$$

where

$$
F_{\text{piston}} = -\frac{\partial^2 \Phi_B}{\partial T^2} + \frac{1}{\Omega^2 \pi^2} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi_B}{\partial X} + \frac{1}{\Omega^2 \pi^2} \frac{\partial^2 \Phi_B}{\partial X^2}
$$
$$
- \frac{2}{\Omega} \frac{\partial^2 \Phi_A}{\partial X \partial T} \frac{\partial \Phi_B}{\partial X} - \frac{2}{\Omega} \frac{\partial^2 \Phi_B}{\partial X \partial T} \frac{\partial \Phi_A}{\partial X} - \frac{2}{\Omega} \frac{\partial^2 \Phi_B}{\partial X \partial T} \frac{\partial \Phi_B}{\partial X}
$$
$$
- \frac{(\gamma-1)}{\Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi_A}{\partial T} \frac{\partial \Phi_B}{\partial X} - \frac{(\gamma-1)}{\Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\}
$$
$$
\times \frac{\partial \Phi_B}{\partial T} \frac{\partial \Phi_A}{\partial X} - \frac{(\gamma-1)}{\Omega} \left\{ \frac{1}{S} \frac{dS}{dX} \right\} \frac{\partial \Phi_B}{\partial T} \frac{\partial \Phi_B}{\partial X}
$$

$$-\frac{(\gamma-1)}{\Omega}\frac{\partial\Phi_A}{\partial T}\frac{\partial^2\Phi_B}{\partial X^2}-\frac{(\gamma-1)}{\Omega}\frac{\partial\Phi_B}{\partial T}\frac{\partial^2\Phi_A}{\partial X^2}$$

$$-\frac{(\gamma-1)}{\Omega}\frac{\partial\Phi_B}{\partial T}\frac{\partial^2\Phi_B}{\partial X^2}-\frac{G_B}{\pi^3\Omega}\left\{\frac{1}{S}\frac{dS}{dX}\right\}\frac{\partial^2\Phi_B}{\partial T\partial X}$$

$$-\frac{G_B}{\pi^3\Omega}\frac{\partial^3\Phi_B}{\partial T\partial X^2}+\frac{G_S}{R\Omega^{1/2}}\frac{\partial\Phi_B}{\partial T}. \tag{11}$$

F_{piston} consists all terms associated with Φ_B relating to the piston motion. Many possible functions can be chosen for Φ_B to satisfy boundary conditions (7). A simple one is taken in the present study,

$$\Phi_B = U^*\frac{X^2}{2}\sin(T). \tag{12}$$

The nonlinear equation (10) for Φ_A, together with the homogeneous boundary conditions (9), are solved by the Galerkin method[4,6] in this study. Once $\Phi=\Phi_A+\Phi_B$ is obtained, the pressure in the resonator can be evaluated by

$$\frac{P}{P_0}=\left[1-(\gamma-1)\pi^2\left(\Omega\frac{\partial\Phi_A}{\partial T}+\Omega\frac{\partial\Phi_B}{\partial T}+\frac{1}{2}\left(\frac{\partial\Phi_A}{\partial X}+\frac{\partial\Phi_B}{\partial X}\right)^2\right.\right.$$

$$-\frac{G_B}{\pi^3}\frac{1}{S}\frac{\partial}{\partial X}\left(S\frac{\partial\Phi_A}{\partial X}+S\frac{\partial\Phi_B}{\partial X}\right)$$

$$\left.\left.+\frac{G_S\Omega^{1/2}}{D}(\Phi_A+\Phi_B)\right)\right]^{\gamma/(\gamma-1)}. \tag{13}$$

The resonance frequency of the resonator can be found in seeking the maximum amplitude of pressure. It can also be estimated from the following formula[6] for an exponentially expanded resonator:

$$k_n^2=\left(\frac{\omega_n}{c_0}\right)^2=\left(\frac{n\pi}{l}\right)^2+\frac{1}{4}\left(\frac{\alpha}{l}\right)^2\quad\text{for }n=1,2,3,\dots, \tag{14}$$

where k_n are the nature acoustic wave numbers of different resonance modes and α is the flare constant. For the fundamental mode ($n=1$), the dimensionless resonance frequency is

$$\Omega=\frac{\omega_1}{\omega_0}=\sqrt{1+\frac{1}{4}\left(\frac{\alpha}{\pi}\right)^2}. \tag{15}$$

B. Calculated results for resonators with piston driving

The cross-sectional area of the resonator is given by

$$S=S_0 e^{\alpha X}, \tag{16}$$

where α is the flare constant indicating the rate of expansion. $\alpha=0$ represents a straight resonator without expansion. The resonators in the following calculation are of the same geometry of $l=0.2$ m, $r_0/l=0.02$, and the gas inside the resonators is air. Figure 2 illustrates the resonance pressure wave forms at the small end ($X=0$) of exponentially expanded resonators with various flare constants. The piston driving strength in this case is $U^*=5\times10^{-4}$. It shows sawtooth wave forms with low pressure amplitudes for the
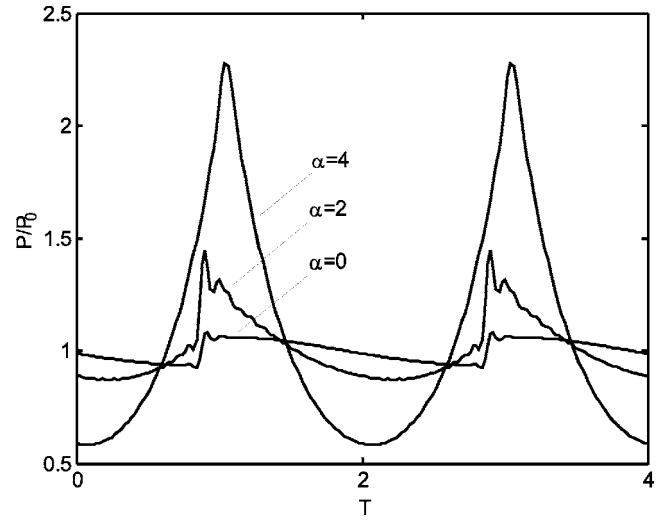
FIG. 2. The wave forms in exponentially expanded resonators with different flare constants. The driving force is $U^*=5\times10^{-4}$.

straight ($\alpha=0$) and slightly expanded ($\alpha=2$) resonators, and shock-free wave forms with high pressure amplitudes for highly expanded ($\alpha=4$) resonator. These are typical characteristics in shaped resonators under shaking.

The resonance frequencies of the resonators with different expansion rates α are identified based on the driving frequencies that produce maximum pressure amplitudes. The results are shown in Fig. 3, together with the frequencies calculated by Eq. (15). It is seen that the resonance frequency basically follows Eq. (15) to increase with the expansion rate. A strong driving ($U^*=5\times10^{-4}$) increases the resonance frequencies compared to the weak driving ($U^*=5\times10^{-6}$), which shows the hardening behavior. The calculated results for $\alpha=5.6$, $\Omega\approx1.35$ are depicted in Fig. 3 for the purpose of comparison with experimental data.

It is known that high amplitude pressures will be obtained at the small end in a shaped resonator. The ratios of
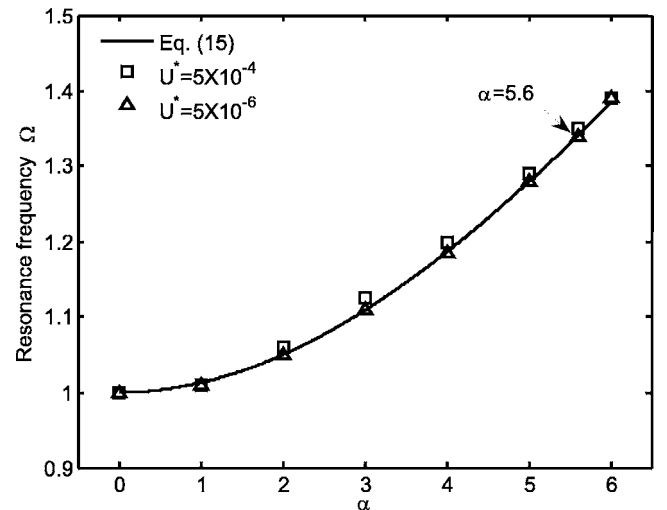
FIG. 3. The dimensionless resonance frequency of exponentially expanded resonators with different flare constants. The open squares and triangles are the results obtained by solving nonlinear wave equations (10) and (13) at strong and weak driving strengths, respectively. The solid line shows the predicted results by using Eq. (15).

Luo *et al.*: Pressure resonance generated by boundary driving   2517
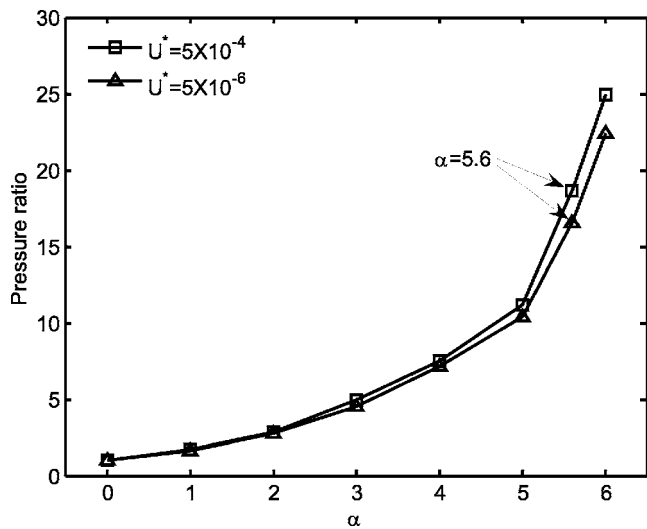
FIG. 4. The ratio of pressure amplitudes at the small end to the big end in exponentially expanded resonators with different flare constants at strong and weak driving forces.

the pressure amplitude at the small end to the pressure amplitude at the big end are calculated for resonators under different expansion rates and driving strengths. The results, plotted in Fig. 4, show that the pressure amplitude ratio increases slowly with the expansion rate in the lower range where the flare constant $\alpha < 5$, and it goes up much faster after $\alpha > 5$, especially under higher strength driving. The pressure amplitude ratio will be used as one of the resonance characteristics for shaped resonators to evaluate the experimental results.

C. Comparison of piston driving with shaking

The piston driving in the present study is compared with the entire shaking on the same resonator shown in Fig. 1. In order to provide a base for the comparison, the strengths in two driving methods must be related. It is assumed that the piston is of the same acceleration as that produced by the shaking on the resonator, i.e., $A^* = U^* \Omega$ (where $A^* = a_0 / l \omega_0^2$).
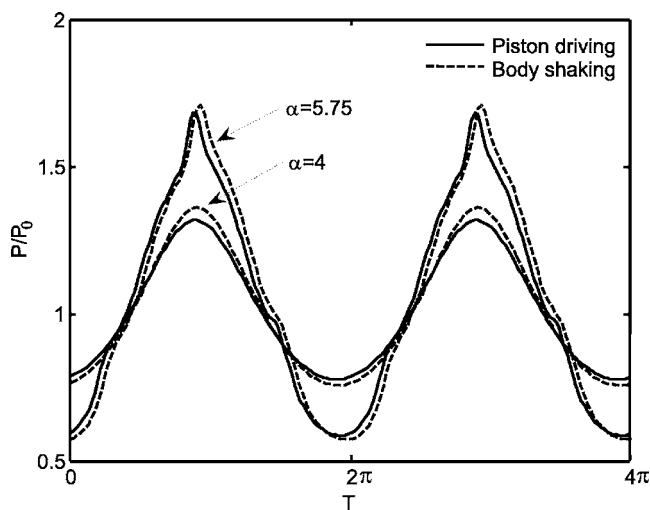
TABLE I. Dimensionless acceleration A for shaking and velocity U^* for piston driving in the force set used in Fig. 6.

Force set	1	2	3	4	5
A'	1×10^{-4}	2×10^{-4}	3×10^{-4}	4×10^{-4}	5×10^{-4}
Ω	1.357	1.361	1.363	1.365	1.367
$U^* = A / \Omega$	7.37×10^{-5}	1.47×10^{-4}	2.20×10^{-4}	2.93×10^{-4}	3.66×10^{-4}

In the case of shaking, the pressures are obtained by solving Eq. (1) under homogeneous boundary conditions at $X = 0$ and $X = 1$. Figure 5 illustrates the pressure wave forms at the small end of resonators with the flare constant $\alpha = 4$ and 5.75, respectively. The dashed lines are the pressures generated by shaking and the solid lines are produced by piston driving. The driving strengths are $A^* = 1 \times 10^{-4}$ in shaking for both resonators. The corresponding piston velocities are $U^* = 8.42 \times 10^{-5}$ and $U^* = 7.37 \times 10^{-5}$ for $\alpha = 4$ and 5.75 resonators, respectively. The piston driving strengths are different for two resonators because of the difference in their resonance frequencies. The results in Fig. 5 show that the piston driving generates similar pressures to that generated by the shaking, in terms of the wave forms and amplitudes. In another comparison, the pressure amplitudes at two ends of a resonator are calculated. The resonator is driven by shaking and piston, respectively, with different strengths. A set of shaking acceleration A^* and the corresponding U^* is denoted as a force set, and five force sets are listed in Table I. The results presented in Fig. 6 show that the piston driving has the same effect as the shaking in generating high amplitude pressures. It should be pointed out that the comparisons conducted here are still qualitative, although the results produced by the two driving methods agree well. This is because the shaking will apply an acceleration uniformly on the gas inside the resonator, while the piston driving only gives the same acceleration to the gas immediately contacting to the piston at $X = 1$.

FIG. 5. The small end pressure wave forms in two exponentially expanded resonators ($\alpha = 4$ and 5.75), obtained by the piston driving and shaking, respectively.
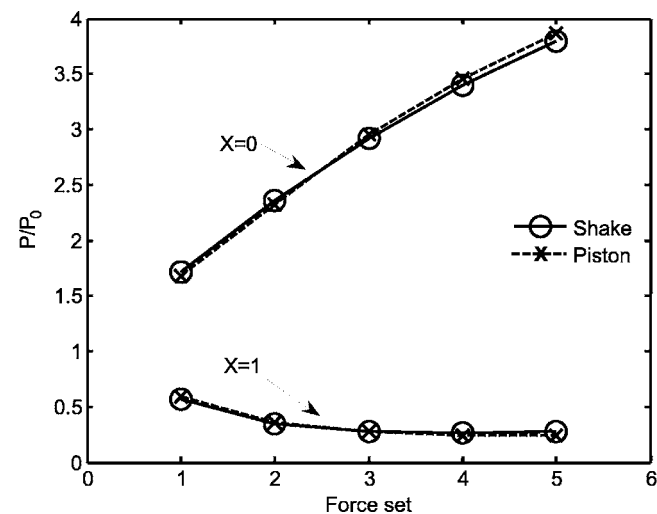
FIG. 6. Comparison of pressures at the small end ($X = 0$) and large end ($X = 1$) produced by shaking and piston driving. The driving strengths associated with the force sets are listed in Table I.

Luo *et al.*: Pressure resonance generated by boundary driving
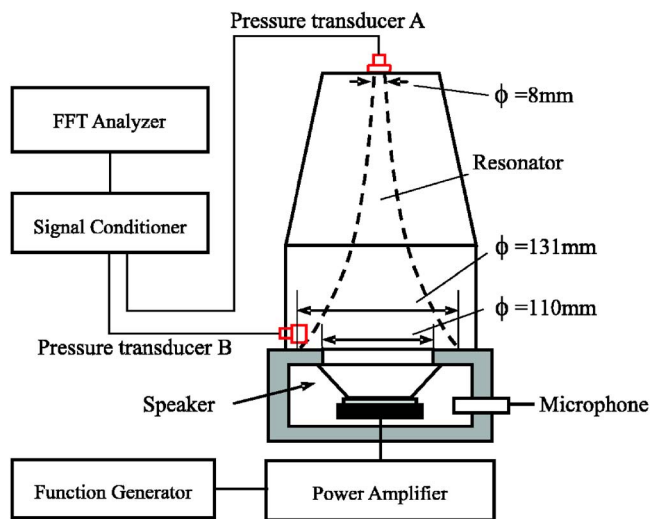
FIG. 7. (Color online) Schematic illustration of the experimental setup for the axisymmetric resonator.

FIG. 8. The frequency responses measured by pressure P_A at different driving strengths, indicated by the microphone outputs. The resonance frequency is found to be 1100 Hz.

III. EXPERIMENT

The experimental study consisted of two parts: (1) An axisymmetric resonator driven by a speaker at one end and (2) a low aspect ratio rectangular resonator driven by a piezoelectric disk at one side of the cavity.

A. Experiment setup for axisymmetric resonator

Figure 7 is a schematic illustration of the experiment setup, including a resonator, a speaker box, two pressure sensors, and other supporting equipment. The resonator was made of aluminum alloy and was exponentially expanded with the flare constant $\alpha = 5.6$. The resonator had a small end of diameter 8 mm and length 200 mm. The small end of the resonator was sealed and the large end was connected to the speaker box through a 110-mm-diam-hole. A 5 in. 40 W loudspeaker was mounted behind the hole to function as the piston. There was a 1/4 in. microphone behind the speaker to measure the sound pressure levels inside the back space of the box. Since the back space of the box was isolated from the resonator by the speaker diaphragm, the sound pressure level measured there was not directly related to the resonance pressures inside the resonator but served as an indication of the piston driving strength. The signals from a function generator (HP 33120A) were amplified by a power amplifier (CROWN D-75), and then were input to the speaker. Two pressure transducers (PCB 103B12) were fixed at the two ends of the resonator to measure dynamic pressures inside the resonator. The pressure transducer at the small end was denoted as A and the pressures measured from it were denoted as P_A accordingly. Similarly, the pressures measured at the big end were denoted as P_B. The signals from the transducers were received by a signal conditioner (PCB 483A) and analyzed by a signal analyzer (ONO SOKKI CF-5220Z).

B. Experimental results of axisymmetric resonator

The resonance frequency of the resonator was identified by measuring the frequency responses of pressure at the small end under various driving strengths. The results are shown in Fig. 8, in which the sound pressure levels (80, 100, and 120 dB) are measured by the microphone to indicate the driving strength. The resonance frequency was found to be 1100 Hz for all three driving strengths, indicating that the driving in the present study was not strong enough to have significant effect on the resonance frequency. The measured resonance frequency $f = 1100$ Hz can be converted to the dimensionless resonance frequency to give $\Omega = 1.29$, which is about 4% less than the calculated value shown in Fig. 3. Considering the speaker used in the experiment is not the flat piston assumed in the modeling, the difference of 4% is quite reasonable. The pressures P_A and P_B obtained in the experiment are used to find the ratio of pressure amplitudes between two ends of the resonator, and the results are presented in Fig. 9. It is found from the experimental data that the pressure ratio between two ends of the resonator is about 17.6, which agrees well with the modeling results 17–19, as
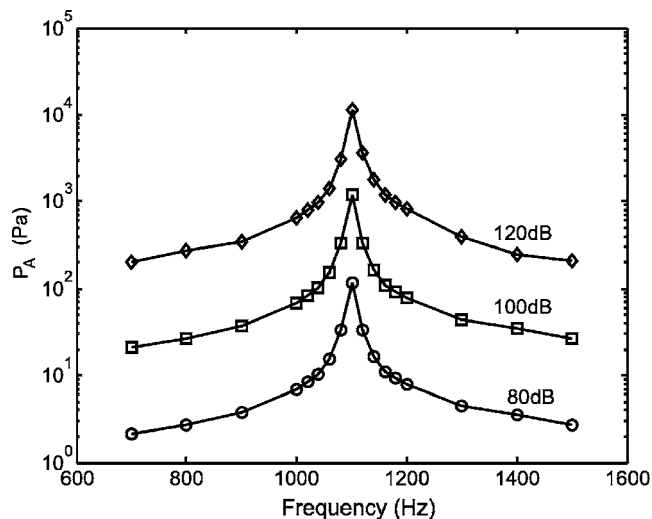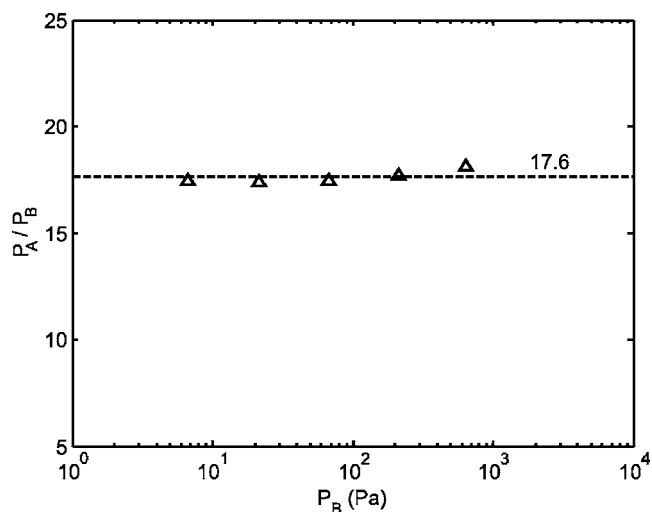
FIG. 9. The ratio of pressure amplitudes at the small end to the big end of the resonator measured from the experiment, indicated by open triangles. The dashed line is the mean value ($P_A/P_B = 17.6$).
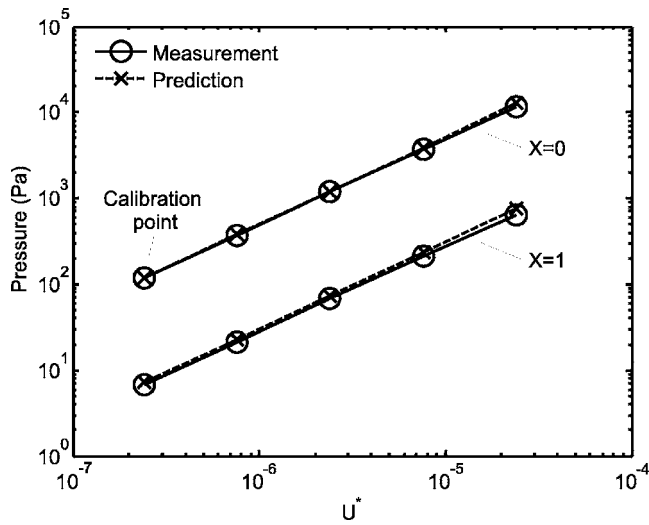
FIG. 10. The pressures at the small end ($X=0$) and the big end ($X=1$) measured in the experiment compared to the calculated values.

indicated in Fig. 4. The results plotted in Fig. 9 show that the pressure amplitude ratios under stronger driving are slightly higher than that under weak driving, which has also been observed in the modeling results shown in Fig. 4.

In order to compare the pressures measured at two ends of the resonator with the calculated values from the modeling, the piston velocity U^* corresponding to the speaker power in the experiment needs to be obtained. This is done based on the fact that the box behind the speaker is sealed, so that the equivalent piston velocity U^* can be assumed to be proportional to the pressure fluctuations behind the speaker, i.e., $U^*=k10^{SPL/20}$, where SPL is the sound pressure level measured by the microphone behind the speaker, k is a constant. In the experiment, ten pressure data from two ends of the resonator were recorded under five different driving SPL values. One datum was used to calibrate for k, so that the piston velocities corresponding to five SPL values were obtained. The calculated pressures under these five piston driving strengths (U^*) were therefore calculated and compared with the experimental data. The results are shown in Fig. 10, in which the predicted pressures at the small end ($X=0$) are compared to the measured values P_A in the experiment, and the predicted pressures at the big end ($X=1$) are compared to P_B measured in the experiment. A good agreement is seen in this comparison.

C. Low aspect ratio rectangular resonator

The boundary driving was further tested by conducting the experiment on a low aspect ratio rectangular resonator. This is to explore the possibility of generating high amplitude dynamic pressures with various driving and resonator configurations, so as to extend their applications, such as in miniature/microfluidic devices. The low aspect ratio configuration is a common feature in miniature fluidic devices fabricated by surface machine technologies. Figure 11 is an assembly drawing of the resonator. The resonator was formed between a cavity plate and a cover plate. The cavity was machined on a brass plate and the depth of the cavity was 3 mm. The profile of the cavity consists of a 3-mm-diam
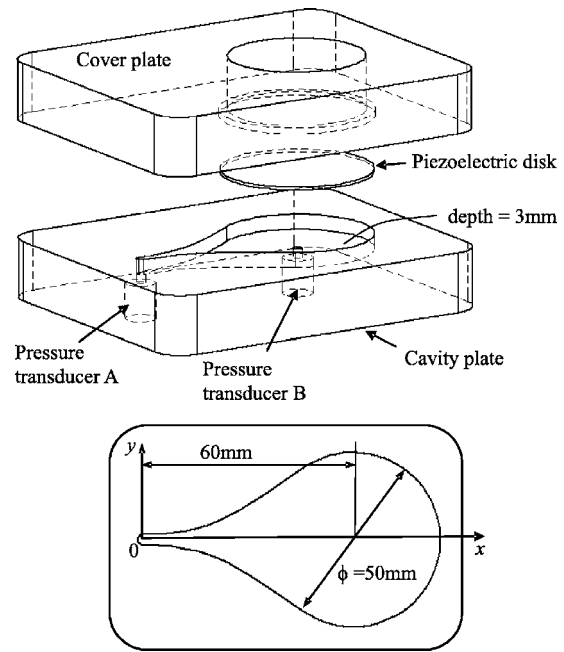
FIG. 11. Schematic illustration of the low aspect ratio rectangular resonator used in the experiment.

semicircle at the small end, a 50-mm-diam semicircle at the big end, and a section linking two semicircles. The profile of the link section is

$$y = \begin{cases} 1.5e^{0.0583x}, & 0 < x \leqslant 42.1 \\ \sqrt{25^2 - (60-x)^2}, & 42.1 < x \leqslant 60. \end{cases} \quad (17)$$

A piezoelectric disk was mounted, as part of the surface, on the cover plate. The diameter of the disk was 50 mm and the thickness was 0.5 mm. The cover plate was fixed onto the cavity plate, and the resonator was formed with the piezoelectric disk being one sidewall of the resonator cavity. The resonator was under the boundary driving when driving voltages were applied on the piezoelectric disk. The pressure transducers were the same as that used for the axisymmetric resonator: One was located below the small end to give pressure P_A and another was located below the center of the semicircle at the big end to give P_B. The resonance pressures P_A and P_B under different driving voltages, from 2 to 23 V, are plotted out in Fig. 12. It shows that the pressures at both ends are linearly increased with the driving voltages applied on the boundary disk, and the pressures at the small end are much higher than that at the big end, which is a typical characteristic observed in other shaped resonators. The ratio of P_A to P_B is calculated based on the data shown in Fig. 12. The average value for the ratio in this case is found to be 20.2, which is comparable to 17.6 obtained in the exponentially expanded resonator with the flare constant $\alpha=5.6$. The resonance frequency for the low aspect ratio rectangular resonator was measured in the experiment to be 4020 Hz. By taking the distance from the small end to the center of the semicircle at the big end as the characteristic length, which is 60 mm shown in Fig. 11, the dimensionless resonance frequency can be calculated to be $\Omega=1.41$. The measured pres-

Luo *et al.*: Pressure resonance generated by boundary driving
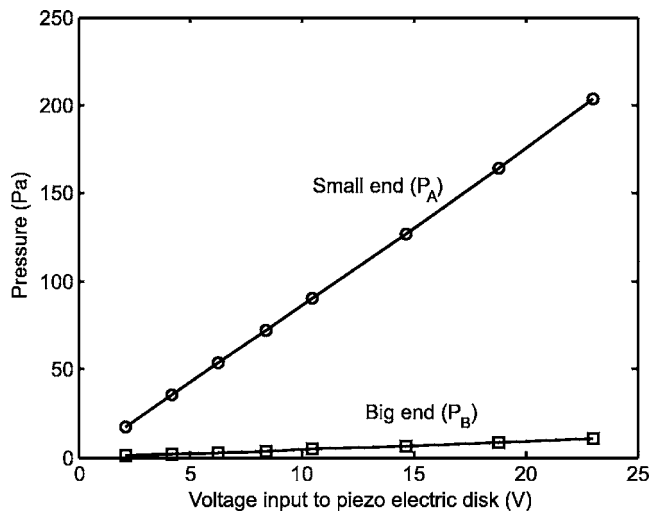
FIG. 12. The pressures obtained from the transducers *A* and *B* in the low aspect ratio rectangular resonator at different driving voltages applied to the piezoelectric disk. The resonance frequency is 4020 Hz and the average pressure ratio is $P_A/P_B = 20.2$.

sure ratio and dimensionless frequency demonstrate that the low aspect ratio rectangular resonator driven by the sidewall has resonance characteristics similar to axisymmetric resonators.

IV. CONCLUSIONS

The generation of high amplitude pressures in shaped resonators by boundary driving has been studied. The configurations of resonators in the study include axisymmetric resonators driven by pistons and a low aspect ratio rectangular resonator driven by a sidewall. In the theoretical modeling, the nonhomogeneous boundary condition associated with the piston driving has been removed by a known function, so that the Galerkin method can be applied. The calculated results, including the resonance frequencies, pressure wave forms, pressures at the two ends of the resonators and their ratios, are compared with the results generated by shaking the resonators and measured from the experiments. The experimental results on the low aspect ratio rectangular resonator with the sidewall driving show the same resonance characteristics as that observed in axisymmetric resonators. The present study demonstrates that the boundary driving, incorporated with properly designed resonators, could be a promising method to provide useful dynamic pressures for various applications, especially in small scale devices.

[1] C. C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, "Measurement of macrosonic standing wave in oscillating closed cavities," J. Acoust. Soc. Am. **104**, 623–636 (1998).
[2] Y. A. Ilinskii, B. Lipkens, T. S. Lucas, T. W. Van Doren, and E. A. Zabolotskaya, "Nonlinear standing waves in an acoustical resonator," J. Acoust. Soc. Am. **104**, 2664–2674 (1998).
[3] Y. A. Ilinskii, B. Lipkens, and E. A. Zabolotskaya, "Energy losses in an acoustical resonator," J. Acoust. Soc. Am. **109**, 1859–1870 (2001).
[4] C. Luo, X. Y. Huang, and N. T. Ngyuen, "Effect of resonator dimensions on nonlinear standing waves," J. Acoust. Soc. Am. **117**, 96–103 (2005).
[5] M. F. Hamilton, Y. A. Ilinskii, and E. A. Zabolotskaya, "Linear and nonlinear frequency shifts in acoustical resonator with varying cross section," J. Acoust. Soc. Am. **110**, 109–119 (2001).
[6] R. R. Erickon and B. T. Zinn, "Modeling of finite amplitude acoustic wave in closed cavities using the Galerkin method," J. Acoust. Soc. Am. **103**, 1863–1870 (2003).
[7] N. Sugimoto, M. Masuda, and T. Hashiguchi, "Frequency response of nonlinear oscillations of air column in a tube with an array of Helmholz resonators," J. Acoust. Soc. Am. **114**, 1772–1784 (2003).
[8] M. Masuda and N. Sugimoto, "Experiments of high-amplitude and shock-free oscillation of air column in a tube with array of Helmholtz resonators," J. Acoust. Soc. Am. **118**, 113–123 (2005).
[9] N. T. Nguyen and S. T. Wereley, *Fundamentals and Applications of Microfluidics* (Artech House, Boston, 2002).
[10] N. T. Nguyen, X. Y. Huang, and K. C. Toh, "MEMS-Micropumps: A review," ASME J. Fluids Eng. **124**, 384–392 (2002).

Comparison of measurements and predictions of sound propagation in a valley-slope configuration in an inhomogeneous atmosphere

T. Van Renterghem[a] and D. Botteldooren
Ghent University, Department of Information Technology, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium

P. Lercher
Medical University of Innsbruck, Section of Social Medicine, Innsbruck, Austria

Mountainous areas form a very specific context for sound propagation: There is a particular ground effect and meteorological conditions are often extreme. In this paper, detailed sound propagation calculations are compared to noise measurements accompanied by meteorological observations. The sound source considered is road traffic along the center axis of a valley. Noise levels were measured in two cross sections, at three locations each: one on the valley floor and two on the slopes, up to 166 m above the source. For the numerical calculations, the rotated Green's function parabolic equation method is used, taking into account the undulation of the terrain and an inhomogeneous atmosphere. Typical parameters of this method were optimized for computational efficiency. Predictions agree with measurements to within 3 dBA up to propagation distances of 1 km, in windless conditions. The calculations further show that the terrain profile is responsible for an increase in sound pressure level at distant, elevated points up to 30 dBA compared to a flat ground situation. Complex temperature profiles account for level changes between -3 dBA and $+10$ dBA relative to a homogeneous atmosphere. This study shows that accurate sound level prediction in a valley-slope configuration requires detailed numerical calculations. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717765]

I. INTRODUCTION

When comparing sound propagation in valley-slope configurations with sound propagation over flat terrain, two main differences are observed.

First, the undulation of the terrain influences sound propagation strongly. At some locations, the receiver is shielded from the source by the terrain. In that case, there is no direct sound path, and sound reaches the receiver only by diffraction and refraction over hills. In other situations, receivers are located on slopes high above the valley floor and have direct visibility of the source. They are exposed to significantly higher sound pressure levels than receivers at the same distance at the level of the valley floor would have. The main reason for this is the reduced ground attenuation. In addition, in concave valleys, multiple reflected sound may converge at the up-slope receiver.[1] In very narrow valleys, sound reflected on the opposite slopes may also contribute significantly to the overall level.[1]

Second, typical meteorological conditions are found in mountainous regions. There is often a large variability of the meteorological parameters in space and time. Besides the influence that mountains exert on the large-scale wind flow (e.g., channeling along the valley axis and the presence of recirculation zones behind orographic obstacles), some typical, thermally driven air currents occur in valley-slope configurations (so-called slope winds). Wind parameters were not measured in sufficient detail during the measurement campaign in the underlying study, and will therefore not be considered in this paper.

Temperature effects can be more prominent in valleys compared to flat terrain. The transition from the (stable) nightly temperature inversion situation to an unstable atmosphere during daytime can happen very quickly, once the sun rays reach the valley floor. The valley orientation plays an important role in this respect and might cause a delay of several hours with regard to the moment of temperature inversion breakup.[2] The width-height ratio of a valley influences both the depth and lifetime of the temperature inversion layer.[3]

In Ref. 4, a meteorological meso-scale model was used to simulate temperature profiles and the development of slope-wind systems in a narrow, two-dimensional valley, in the absence of large-scale winds. This information was then used by a numerical sound particle model. This study yielded large variations of sound levels during the course of a day because of the state of the atmosphere. No measurement campaign was set up to check their findings.

In this paper, simultaneous noise measurements and meteorological measurements were performed in order to check sound propagation calculations. The situation of interest is upslope sound propagation, orthogonal to the valley axis. This corresponds to a typical situation in valley-slope con-

[a] Author to whom correspondence should be addressed. Electronic mail: timothy.van.renterghem@intec.ugent.be
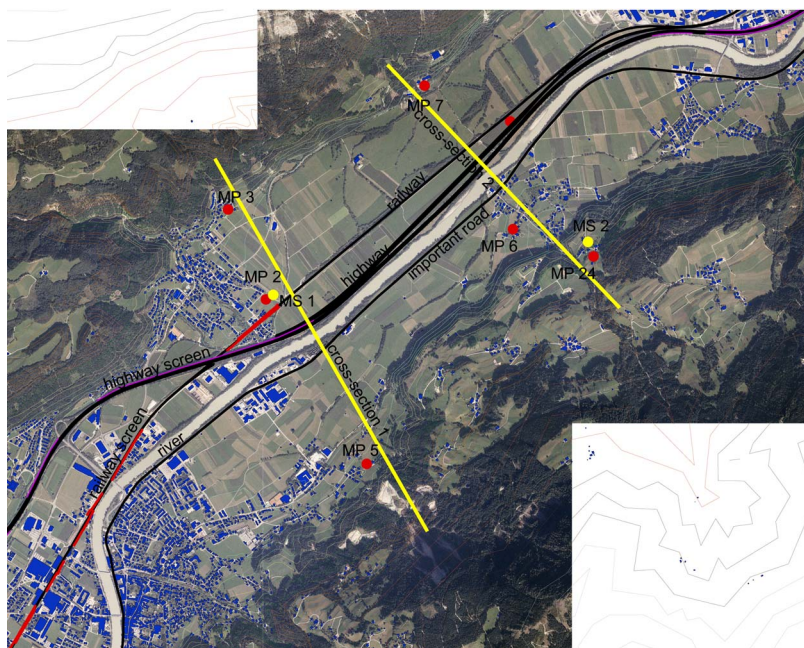
FIG. 1. (Color online) Orthophoto of the valley under consideration. The following items are indicated: the microphone positions (MP), the locations of the meteo stations (MS), the highway, the main road, the railway, the river, the locations of highway screens and railway screens, and the buildings. Height contours are shown as well.

figurations, where a highway or railway follows the valley axis and where dwellings are present on the slopes. A detailed comparison between noise measurements and (wave-based) sound propagation calculations in such situations has not been reported previously. Focus is on the effect of an inhomogeneous atmosphere. Large-scale wind systems will influence sound propagation to a lesser extent in this—typical—configuration since wind direction is usually aligned with the valley axis. Slope winds are directed orthogonal to the valley axis, but are not considered here.

This paper is organized as follows. In Sec. II, the region where the study was performed is described. In Sec. III, the acoustical and meteorological measurement setup is described. It is also indicated how the data was preprocessed. Details on the numerical model and the calculation methodology can be found in Sec. IV and Appendix. A comparison between measurements and predictions is made in Sec. V. In Sec. VI, conclusions are drawn.

II. SITE DESCRIPTION

The Unterinntal region is located in the Alps, in the western part of Austria. The measurements were performed in a 2-km-wide valley in the district Schwaz. An orthophoto of the area under consideration is shown in Fig. 1. The valley floor is located at an elevation between 530 and 540 m. The valley is oriented in general from North-East to South-West. The mountains surrounding the valley have ridges of over 2000 m in height. In the center of the valley, there is a highway (Inntal Autobahn, A12), an important main road (Tiroler Bundesstrasse, B171), and a railway close to the highway.

In this paper, focus is on sound propagation from road sources to a number of locations on the slopes of the valley, at lower elevation. The difference in elevation between the road and the microphone positions is at most 166 m. The presence of dwellings and thus possibly noise annoyed people at higher altitude on the slope is limited.

Two cross sections were selected and in each of them three microphones were placed. The microphone positions in each cross section lie more or less on a straight line, orthogonal to the roads. Near cross section 1, the highway (A12) is the dominant noise source. Near cross section 2 a busy road (B171) dominates the noise climate. An overview of the orthogonal distances to the relevant roads, the ground elevation at the microphone locations, as well as the elevation of the roads near each microphone can be found in Table I. All microphones, with the exception of microphone 6, are placed 2 m above the ground. Microphone 6 is placed on the roof of a small building for practical reasons. A portable meteo station is used to gather on-site meteorological data.

Figure 1 shows a map of the area clearly indicating the roads, railway, noise barriers along highway and railway, houses and other buildings, and the measurement locations. Sound measurement stations are identified by the microphone numbers used in Table I.

III. MEASUREMENTS AND PROCESSING

The measurement campaign lasted from November 2005 till January 2006. Simultaneous noise measurements and meteorological observations were performed.

TABLE I. Overview of microphone positions during the measurement campaign.

Cross-section ID	Microphone ID	Orthogonal distance towards dominant road noise source (m)	Elevation of the roads (m) near microphone	Elevation at receiver (m)
1	2	331	530–541	540
1	3	1188	530–560	583
1	5	1216	530–563	579
2	6	542	530–540	541
2	7	796	530–540	542
2	24	1153	530–541	696

TABLE II. Overview of clustered datasets.

Cross-section ID	Cluster ID (temperature profile ID)	Number of locations	Number of retained measurements in cluster	Additional meteo information in cross section	Measurement period: day(s)	Measurement period: hours
1	1	3	53	Yes (MS 1)	20/11/05	16.00–19.00 h
2	2	3	148	No (MS 1 is used)	26/11/05–01/12/05	9.00–16.00 h
2	3	3	67	No (MS 1 is used)	27/11/05	11.00–13.00 h
2	4	2	59	Yes (MS 2)	20/12/05–21/12/05	23.00–08.00 h

A. Meteo measurements and processing

A Vaisala MAWS201 Automatic Weather Station was used to gather basic meteorological information in each cross section. The data consist of air temperature, relative humidity, atmospheric pressure, wind speed, and wind direction, all measured at a single height. Every minute, this data was logged. The wind speed was measured with an anemometer at 2 m above the ground. This information is insufficient to estimate wind speed profiles. It nevertheless allows one to exclude 1 min periods where wind might influence sound propagation. At the same time, selecting measurements at low wind speeds only also prevents microphone induced wind noise. A maximum value of the wind speed equal to 0.5 m/s was used. To exclude measurements made during precipitation, only those measurements made when relative humidity was lower than 80% were retained. As a result, the comparison between prediction and measurement will be limited to dry, windless conditions.

Temperature profiles were obtained by means of eight ventilated temperature sensors attached to the posts of a cable way (the "Kellerjoch Bahn"). The heights of the sensors ranged from 540 m (valley floor) to 1341 m. The Kellerjoch Bahn is located North-West from cross section 1 and cross section 2, at 4.5 and at 6.5 km, respectively. Every 15 min, the temperature at all heights was logged simultaneously. Thus, air temperature is known along a single line on a slope. In an ideal situation, temperature should be measured at a fixed location (e.g., in the center of the valley) at different heights but this was not practically achievable.

B. Noise measurements and processing

In each cross section, simultaneous measurements were performed with Svantek 1/2 in. SV22 condenser microphones. Overall equivalent A-weighted sound pressure levels are stored every second ($L_{Aeq,1\ s}$).

Besides the main road noise that is of interest for this study, noise from train passages and local events like the passing of a car on a nearby, small road are also recorded by the microphones. To eliminate these disturbances, the raw $L_{Aeq,1\ s}$ measurements are pre-processed based on the hypothesis that the main road noise under study constitutes the constant part of the sound level. The following rule was used: The sound pressure level at a given second is considered to be an event if it is higher than the median noise level in a time window of 5 min, centered on the second under consideration, plus 5 dBA. After events were removed in this way, sound pressure levels were summed to 1 min equivalent sound pressure levels. Only if at least 40 s of measurements remain after event canceling, the 1 min equivalent sound pressure level is kept for further analyses.

C. Combined dataset

The measurement campaign led to a dataset of combined noise and meteorological data. A clustering of these data was done based on (similar) range-independent temperature profiles, in order to limit the number of calculations. The minimum number of data points in each cluster was set to 50 to allow one to draw statistically stable conclusions. The parameters used for clustering were the gradients in air temperature between successive sensors, up to a height of 767 m. Gradients at larger heights did not influence sound propagation over the distances and height differences considered. The (absolute) value of temperature may differ within a cluster. This is acceptable since the gradient in temperature is responsible for refraction of sound.

The preprocessing of the dataset to retain windless periods without precipitation, together with the demand that at least 50 measurements are characterized by similar air temperature profiles, resulted in a drastic decrease of available combined noise and meteo data. The 4270 measurements (available after removing events) in cross section 1 resulted in 53 usable combined data records. For cross section 2, 21,601 measurements gave 274 combined data records, split up into three temperature profiles clusters.

In Table II, the number of (non-successive) 1 min combined data records are shown for the different clusters considered. During all of these selected periods, the ground was snow-covered.

The data for cross section 1 was measured on a single day, between 16.00 and 19.00 h. For cluster 2, data come from different days and were recorded during day hours. The data in cluster 3 come again from a single day, around noon. Cluster 4 contains mainly observations during night hours.
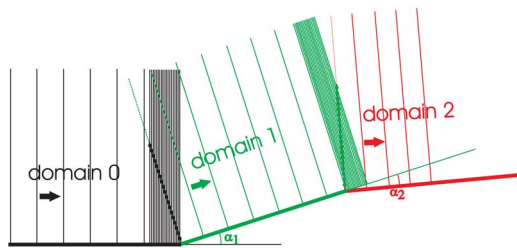
FIG. 2. (Color online) Schematic representation of sound propagation with the GFrPE method in the case of concave and convex ground surface transitions. Three successive flat domains are shown. The vertical lines represent the positions where a column of pressures is calculated. In the transition zones, a large number of propagation steps are needed; α_i represent the difference in slope angle between successive domains.

IV. NUMERICAL PREDICTIONS

A. GFrPE method

A two-dimensional Green's Function Parabolic Equation (GFPE)[5,6] method with a rotated reference frame (abbreviated as GFrPE) is used for the numerical predictions. In this model, the undulating terrain is approximated by a succession of flat domains with different slopes. In each of them, ordinary GFPE calculations are performed. As shown in Fig. 2, the sound field calculation in each domain starts from an array of pressure values, orthogonal to the local slope. The starting field for domain $n+1$ is constructed based solely on calculations in domain n, which is in line with the progressive character of the PE method. When there is a change in the slope angle between successive domains, a number of reduced propagation steps are needed in domain n near the transition to the next domain, to allow one to obtain the pressures at the correct heights for constructing the starting field for domain $n+1$.

This methodology is illustrated in Fig. 2, for both a concave transition (from domain 0 to 1) and a convex transition (from domain 1 to 2). In the case of a convex ground surface, PE calculations are necessary along the (virtual) continuation of the ground in domain n for calculating the pressures of the starting field of domain $n+1$.

This modification to the GFPE method was proposed in Refs. 7 and 8. The GFrPE method has the same benefits as GFPE:

- The computational cost of GFPE is based mainly on the efficiency of the fast Fourier transform (FFT) algorithm. Very fast FFT algorithms are available.
- GFPE allows one to use large step sizes in horizontal direction, that are limited by the inhomogeneity of the atmosphere rather than by the sound wavelength (λ). The maximum acceptable range step varies roughly between 5λ and 50λ in a refracting atmosphere.[6] Since we are aiming at distances up to a few kilometers from the source and for typical traffic noise (including relatively high frequencies), the use of large step sizes is an important advantage. Section IV B discusses how this advantage could be jeopardized by the use of GFrPE.
- Refraction of sound by arbitrary sound speed profiles can be modeled: The sound speed profiles may contain upward and downward refracting parts.
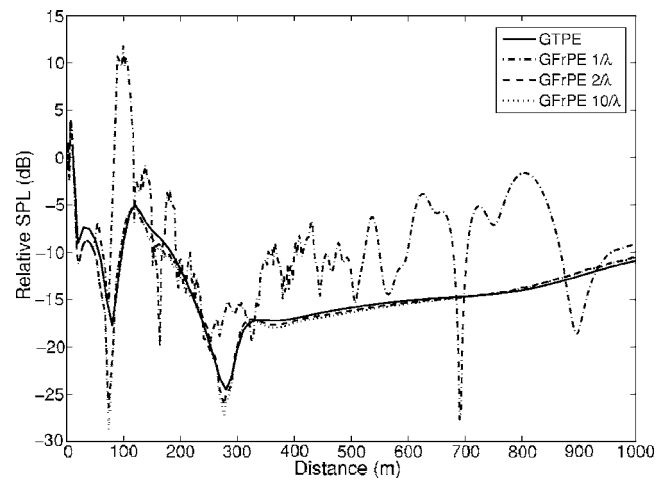
FIG. 3. Effect of the number of calculations per wavelength (n/λ) on the starting fields for GFrPE calculations. See the caption of Fig. 11 for more details on the calculation parameters. The GTPE calculation is used as a reference solution. When two points per wavelength or more are calculated, accuracy is not improved anymore.

- Locally reacting, range-dependent impedance planes can be used to model reflection from the ground.
- Diffraction near hard, infinitely thin screens can be modeled with the Kirchoff approach.[9]

The method was validated for typical road embankment configurations by comparing it with results obtained using the boundary element method.[8] In the Appendix, the GFrPE code is validated for the case of the smooth hill presented in Ref. 10.

B. Improving computational efficiency of GFrPE

The construction of the starting field for domain $n+1$ from field calculations in domain n is computationally very costly. As shown in Fig. 2, a large number of very small propagation steps are needed in the transition zone. A typical starting field would be described by about ten values per wavelength in the vertical direction. In practice, it is not necessary to accurately calculate the field with this resolution on the basis of the propagation in domain n. Instead, linear interpolation in both amplitude and phase can be used to construct the starting field from a lower number of known values. In Fig. 3, the relative sound pressure level as a function of distance for the validation case, described in the Appendix, is shown, with a decreasing number of explicitly calculated points per wavelength on the starting fields. Once 2 points per wavelength (indicated as $2/\lambda$) or more are calculated, the result converges.

GFrPE calculations can also be accelerated by truncating the height of the calculated starting field below the top of the computational grid (or up to the beginning of the absorbing layer). The starting pressure field above this truncation height is obtained by linear extrapolation of the phase angle of the pressure and linear tapering to zero of the magnitude of the pressure as was proposed in Ref. 11. The maximum height still containing relevant information depends on the propagation distance to be covered by the PE model. Figure 4 shows that the relative sound pressure level converges for
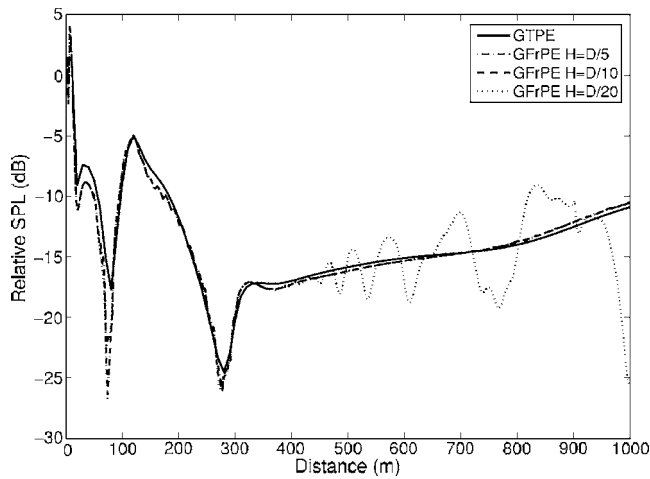
FIG. 4. Effect of the ratio between the maximum height on the starting fields taken into account (H) and the distance between source and receiver (D). See the caption of Fig. 11 for more details on the calculation parameters. The GTPE calculation is used as a reference solution. The use of values of H larger than D/10 only increase the computational cost, while accuracy is not improved anymore.

the test case of the Appendix when the ratio between the propagation distance to be covered (D) and the truncation height of the starting field (H) is at least 10. Decreasing this ratio only increases the computational cost, while there is no gain in accuracy.

C. Comparing relative levels between predictions and measurements

Detailed traffic counts, traffic composition, and traffic speed distribution were not available for the motorways during the measurement campaign. This lack of information can be circumvented by validating the numerical results on relative rather than absolute measurements. In each cross section, the location closest to the road was chosen as the reference measurement: microphone with ID 2 in cross section 1, and microphone with ID 6 in cross section 2. Sound pressure levels at the distant points relative to the reference points were compared with numerical calculations.

The number of vehicles on the road will not influence the comparison of relative sound levels as proposed in the previous paragraph if it can be assumed that the flow is homogeneous over a sufficiently long stretch of the road. This does not hold for vehicle type and vehicle speed. Each combination of vehicle type and vehicle speed results in a typical source spectrum. Keeping in mind the difference in distance between reference point and the more distant points, the contributions of different frequency bands to the total sound pressure level will change. As a result, the relative sound pressure levels will vary. To account for this, the spread in sound pressure level caused by different vehicle types, driving at typical velocities, will be predicted.

D. Calculation methodology for the valley-slope configuration

The methodology for calculating the sound pressure levels from traffic noise at distant receivers uses a number of aspects from the HARMONOISE reference model.[12]

First the three-dimensional propagation problem is split up into a set of two-dimensional problems by subdividing the roads in a number of point sources along its axes.

In a second step, the terrain profile along the line between each point source location and the receiver is extracted from the digital terrain map. In GFrPE, the terrain is approximated as a succession of flat segments, each with a different slope.

The points of intersection between a source-receiver line and obstacles (buildings, highway screens and railway screens) are determined. The GFrPE method is able to handle diffraction over thin, hard screens using the Kirchoff approach: The acoustic field is set to zero on the barrier, and then propagated in forward direction. For simplicity, a building is approximated by such a thin, hard screen located in the center of the building. In this two-dimensional approach, only diffraction over the top of the obstacles is accounted for. It is also assumed that the obstacles are rotated in such a way that their main axis becomes perpendicular to the source-receiver line. These same assumptions were made in the HARMONOISE reference model[12] (except when a three-dimensional ray model is used).

In a third step, the transmission loss from each point source to the receiver is calculated using GFrPE. Temperature profiles are available on a single line along a slope. For simplicity, horizontal temperature stratification throughout the valley is assumed: Only the elevation determines the air temperature. The validity of this approach could be questioned, but it remains the best possible approach with the available meteorological data.

The highest temperature sensor in our case was located at a height of 1341 m (or 801 m above the valley floor). This large span in heights of temperature observations is more than sufficient to accurately model refraction of sound, keeping in mind the distances between sources and receivers. The temperature and sound speed at each position in the vertical PE grid is obtained by linear interpolation between measured data.

Atmospheric attenuation is not included in the GFrPE model, and is therefore added afterwards, using the product αr, where α is the absorption coefficient which is calculated following ISO 9613-1[13] and r is the distance traveled by the direct sound ray between source and receiver. All combinations of air temperature, relative humidity and air pressure that are present in each temperature profile cluster are considered.

In the case of an upward refracting atmosphere, turbulent scattering into the acoustic shadow zone that is formed becomes important. Neglecting this effect often results in unrealistically large attenuations. A standard approach to account for turbulent scattering consists in calculating sound propagation through a number of turbulent realizations of the atmosphere (see, e.g., Ref. 14). The ensemble average of all these realizations follows statistical laws. To have statistically stable results however, at least 50 realizations of the turbulent atmosphere need to be taken into account. As a result, computing times increase dramatically.

Based on experiments, it was found that the sound pressure level relative to free field propagation stays more or less

constant in the acoustical shadow zone formed by an upward refracting atmosphere.[15,16] This constant value depends on the geometry of the problem and on the strength of the turbulence. A value of −20 dB is common in acoustical literature.[10] Truncating the sound pressure level relative to free field comes at no additional computational cost and is therefore preferred for the large scale problem considered in our work. It was observed that this approach led to acceptable results in situations where turbulent scattering into shadow zones becomes important (see further).

In a fourth step, the sound pressure level at the receivers is calculated by choosing an appropriate traffic source spectrum.

Finally, the contributions from all two-dimensional cross sections are added incoherently to find the total sound pressure level at the receiver.

This approach assumes that in each point source, the same type of vehicle is present driving at the same speed. When multiple vehicle types and vehicle speeds are considered, step four (and five) are repeated with an appropriate source spectrum.

E. Model parameters

The numerical parameters were chosen as a compromise between numerical accuracy and computational efficiency.

1. Frequency range

One-third octave bands ranging from 50 to 2500 Hz were considered, covering sufficiently the frequency spectrum of road traffic. Propagation calculations were performed for a single frequency per one-third octave band. The calculations were repeated in the first cross section using three frequencies per one-third octave band for validation. This resulted in differences in total A-weighted sound pressure levels that were less than 0.3 dBA, while the computational cost was three times higher. Thus it was decided to use a single frequency per one-third octave band in all further predictions.

2. Source and receiver heights

A source height above the road surface of 0.5 m was chosen to represent an average over all possible physical noise sources at different speeds for different types of vehicles. The HARMONOISE reference model [12] suggests using different source heights for each source mechanism (rolling noise, engine noise) and each class of vehicles (cars, light trucks and heavy trucks). Since each source height would result in a new propagation calculation, this suggestion was not followed here. The receiver height at all locations was 2 m, except for the microphone with ID 6, which was placed 2 m above the roof of a small building.

3. Discretization of the roads

The stretch of road considered to contribute to the overall noise level was limited by the furthest points being at a distance from the receiver equal to three times the orthogonal distance between road and receiver. In the case of direct sound and in a still and homogeneous atmosphere, a source

at this distance has a contribution to the overall sound level that is nearly 10 dB below the largest contribution. The discretization distance between successive source points along the road axis was first chosen to be 200 m. The number of source points that were used for the different microphone positions ranges from 9 to 35 with such a discretization along the road and using such a marking off. This rough sampling of point sources along the road revealed the zone with the most important contributions, which was in most cases centered around the source point with the shortest distance to the receiver. The stretch of road between the first and the last source point resulting in a sound pressure level of 10 dBA below the most contributing source was subject to further refinement. Additional source points were placed every 100 m. If the total A-weighted sound pressure level obtained using the refined road discretization deviated less than 1 dBA from that obtained using the rougher discretization, convergence was assumed. If not, the refinement procedure was repeated. In most situations a single refinement proved sufficient.

4. Discretization of sound paths

Along each sound path between the source and receiver, the terrain is approximated by flat domains with a length of 100 m. This approximation is acceptable because the relief is reasonably smooth up to the microphone positions. Taking smaller segments largely increases the computational cost mainly because of the smaller spatial step and corresponding larger number of calculations needed in each transition zone.

5. Obstacles

The railway screens in the area under consideration have a height of 3 m. The height of the highway noise barriers ranges from 3 to 4.5 m. The height of individual buildings and houses is not known. An average height of 5 m is chosen. All obstacles are considered to be rigid.

6. Ground modeling

A range-dependent ground impedance is used. The ground directly under the source points is assumed rigid. The river, which is close to the roads, is modeled as a rigid surface as well. All remaining grounds in the source-receiver lines are assumed to be covered with snow. The soil in the buildup areas is not considered separately.

Sound propagation over snow-covered ground has been investigated in detail in acoustical literature.[17,18] Information on, e.g., the thickness of the snow layer or its state (old snow, fresh snow, the degree of compaction, etc.) was not available during the measurement campaign. Therefore, the general-purpose one-parameter ground impedance model of Delany and Bazley[19] was used, with a measured flow resistivity for snow equal to 30 kPa s/m^2 (Ref. 12).

7. Traffic source spectrum

The NORD 2000 traffic source spectra [20] are used. The range of differences in total A-weighted sound pressure level between the distant receivers and the reference receiver are calculated for passenger cars driving at 70, 90, 110, and

130 km/h, for medium-heavy vehicles driving at 70, 90, and 110 km/h, and for heavy vehicles driving at 50, 70, and 90 km/h.

8. GFrPE parameters

A standard Gaussian starting function [5,6] is used to initiate the GFrPE calculations. The absorbing layer as described in the Appendix is applied. The maximum horizontal propagation step is 10λ, and is used as much as possible. Variable horizontal propagation steps are used in order to exactly account for the locations of screens, houses, receivers and impedance changes.

The optimization of characteristic parameters for the GFrPE method was discussed in Sec. IV B. To construct a new starting field for the next domain, two calculations per wavelength are performed in vertical direction followed by linear interpolation to obtain a vertical discretization of 0.1λ. The maximum height of the starting field that is calculated was one tenth of the distance between source and receiver. The starting field above this maximum height is found by extrapolation, as described in Sec. IV B.

A change in slope inclination between successive domains of less than $1°$ is ignored. The starting field for domain $n+1$ is in that case the vertical array of pressures at the border of domains n and $n+1$.

9. Parallel computations

The proposed calculation methodology is computationally costly. This holds especially for the highest frequencies considered, for the oblique (and thus long) sound paths and for the sound paths with a large number of differences in slope inclination between successive domains. In the proposed methodology, calculations are easy to parallelize. Either different sound paths were calculated on different CPU's, or different frequencies were calculated on different CPU's.

V. COMPARISON BETWEEN MEASUREMENTS AND PREDICTIONS

An overall comparison between measurements and predictions is shown in Figs. 5(a)–5(d). The measurements are presented by means of so-called boxplots. The (middle) horizontal line in the box indicates the median of the data. The box is closed by the first and third quartile. The whiskers extend to 1.5 times the interquartile distance above the maximum value inside the box, and to 1.5 times the interquartile distance below the minimum value inside the box. Data points that fall outside the whiskers are considered to be outliers, and are indicated with the plus signs.

Three sets of predicted relative noise levels are included in the plots, indicated by different symbols. The first set is the best available prediction. To gain understanding in the significance of the terrain elevation and temperature effects, additional calculations were performed. First, a homogeneous atmosphere was assumed in the presence of the actual relief. Second, a flat terrain is assumed in a homogeneous

atmosphere. In both calculations all other numerical and geometrical parameters (locations of houses and screens, ground impedance, etc.) remain unchanged.

As discussed in Sec. III C, the combined noise-meteo measurements were clustered based on the temperature profiles. This still leaves an important variance in the relative measurements. The reasons for this variance are summarized below:

- Traffic composition and vehicle speed may be different during the non-successive minutes considered, resulting in different source spectra.
- In each cluster, there is a variation in the combination of air temperature, relative humidity and atmospheric pressure, which is responsible for the magnitude of the atmospheric absorption. An additional clustering on this data would lead to too few cases per cluster.
- An average, clustered temperature profile is chosen for the calculations [see full lines in the temperature profile plots in Figs. 5(a)–5(d)]. Within each cluster, temperature profiles might be slightly different [see dashed lines in the temperature profile plots in Figs. 5(a)–5(d)].

The first two causes of variance in the relative sound pressure levels are accounted for (see Sec. IV C and IV D) and result in the spread in calculation results that can be observed in Figs. 5(a)–5(d). The last one is not included for reasons of computational cost.

The agreement between measurements and numerical predictions is good. The average of the calculations lies close to the median of the measurements. Differences range up to $2–3$ dBA. The spread in the calculations is in most situations very similar to the measured one. This gives confidence in the followed approach.

A comparison between the flat terrain calculations and the calculations using the actual relief, both in a homogeneous atmosphere, indicates that the presence of the sloping terrain is responsible for an increase in sound pressure level for the distant observation points of at least 5 dBA and at most 30 dBA. Comparing the calculations in the case of a homogeneous and an inhomogeneous atmosphere, both using the actual relief, reveals that temperature gradients in this Alpine valley result in a change of sound pressure level ranging from -3 dBA to $+10$ dBA between the distant and reference receiver.

The temperature profiles in the dataset are complex: They contain upward and downward refracting parts, depending on the height. Temperature profiles 1 (see Fig. 5(a)) and 4 (see Fig. 5(d)) are mainly upward refracting. Very close to the ground, a thin temperature inversion layer is observed. This holds also for temperature profile 2 (see Fig. 5(b)); in addition, starting from about 100 m above the valley floor, a temperature inversion layer is observed. Temperature profile 3 (see Fig. 5(c)) is characterized by a well-mixed, unstable layer starting from the ground surface, capped by a strong inversion layer, starting at the same height as in temperature profile 2.

In the first cross section, the relief increases the sound pressure levels at the distant points 3 and 5, relative to the reference point 2. This is shown in Fig. 5(a). When the tem-
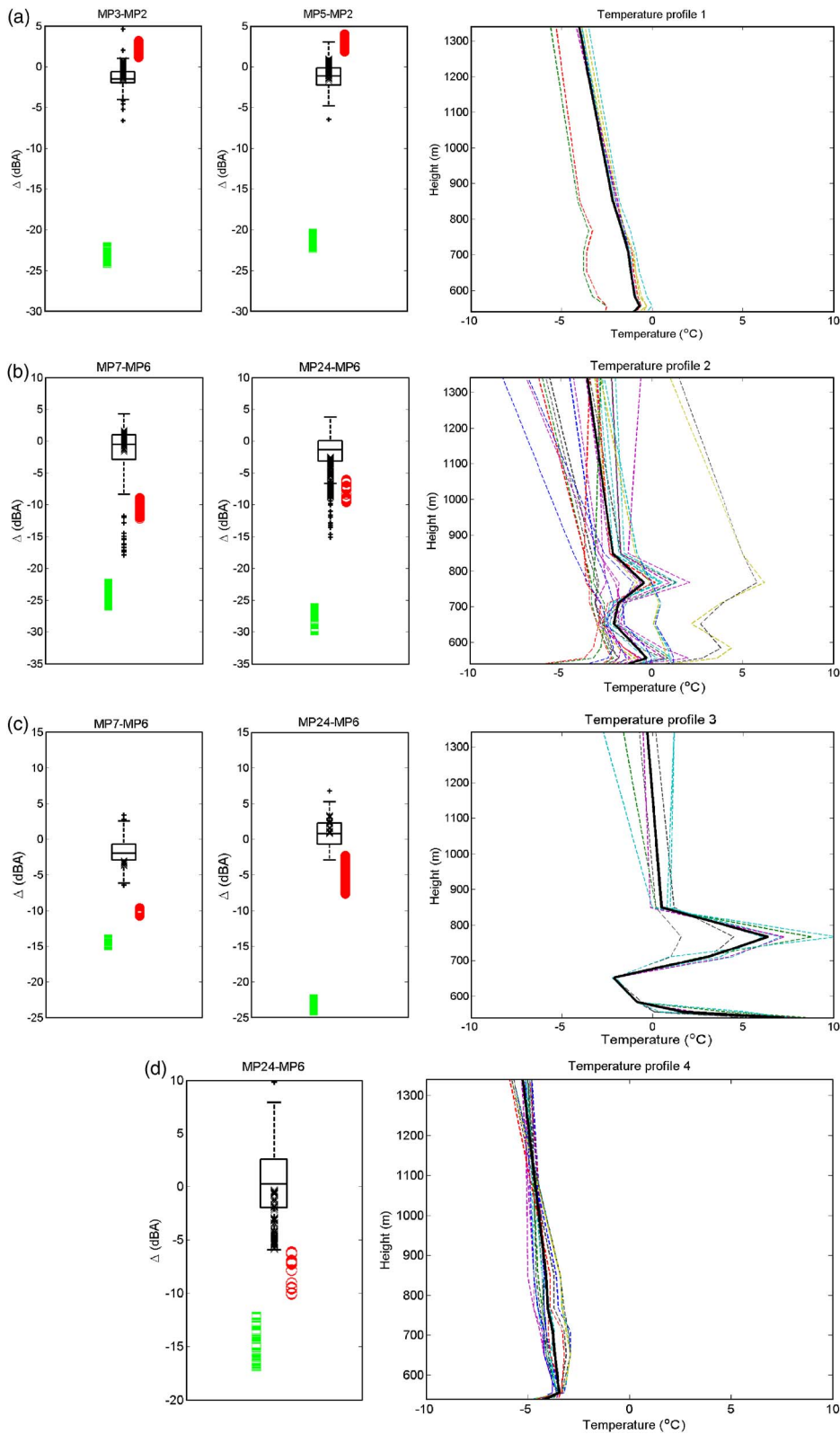
FIG. 5. (Color online) The figures on the left show a comparison between measurements and calculations. The measurements are represented by means of boxplots. $+ -$ signs indicate outliers of the measurements. The series of $x -$ signs indicate the calculations accounting for the actual relief and the actual temperature profile (that can be compared to the measurements). The series of $\bigcirc -$ signs indicate the calculations with the actual relief in a homogeneous atmosphere, the series of $\square$-signs indicate the calculations with a flat terrain in a homogeneous atmosphere. The figures on the right show the temperature profiles forming the clusters (dashed lines), together with the average temperature profiles that were used for the calculations (thick full lines). In part (a), cross section 1 is considered. The sound pressure levels at measuring points (MP) 3 and 5 are shown, relative to the sound pressure level at MP 2 (closest to the road). In parts (b), (c), and (d), cross section 2 is considered. In parts (b) and (c), the sound pressure levels at MP 7 and MP 24 are shown relative to MP 6, for two temperature profile clusters. In part (d), the sound pressure level at MP 24 relative to MP 6 is shown, for temperature profile cluster 4.

perature profile is included, a shift in the other direction is observed. The sound pressure level at microphone 2 slightly increases because of the thin temperature inversion layer very close to the ground. Microphones 3 and 5 receive less sound, because of the upward refracting atmosphere. As a result, the difference in sound pressure level, relative to the homogeneous atmosphere (with relief), decreases, and a relative sound pressure level near 0 dBA is found.

In cross section 2 (see Figs. 5(b)–5(d)), the effects of both the elevation of the terrain and the inhomogeneous atmosphere tend to decrease the difference in sound pressure level between the reference point and the more distant points. Qualitative analysis of what is happening is difficult because of the complex interaction of the mechanisms involved. The terrain profiles corresponding to the different source-receiver lines for microphone 24 are complex, and are
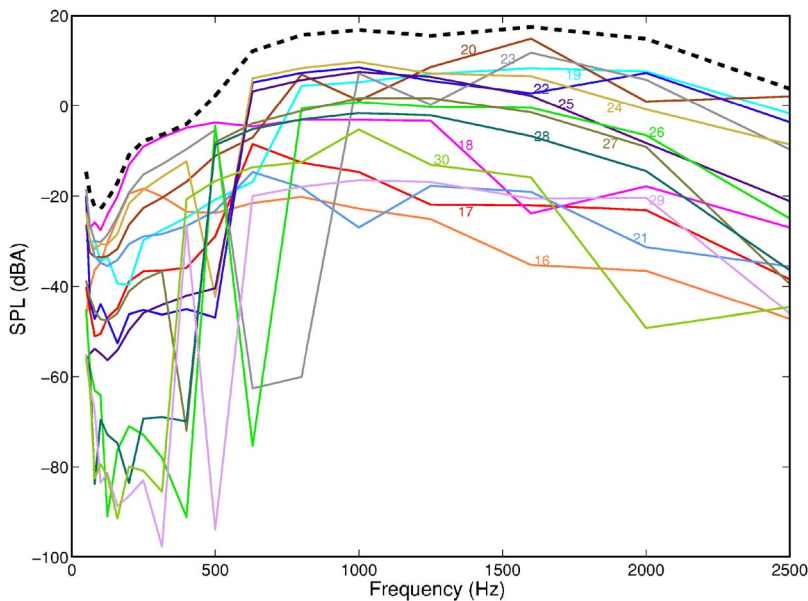
FIG. 6. (Color online) Calculated frequency spectrum at receiver 3 for the 15 most contributing sound paths. A passenger car at 110 km/h is assumed in each source point. The temperature is 0 °C, the relative humidity is 65%. Temperature profile 1 is used. The total A-weighted spectrum is indicated by the thick, dashed line. The magnitudes of the sound pressure levels are arbitrary. Source points after the refinement are not shown. For each source point, the frequency spectrum is clearly different.

characterized by a sudden increase in the elevation of the ground, close to the receiver. These profiles further contain successions of pronounced concave and convex parts. The strongly upward refracting part in temperature profile 3 (see Fig. 5(c)) results in an acoustic shadow zone. Limiting the attenuation in this situation is important to account for turbulent scattering (see Sec. IV D). The situation is further complicated because microphone 6 was placed on the roof of a building. This receiver was (slightly) shielded by the edge of the roof, and this intensifies differences in sound pressure level by the different states of the atmosphere. In contrast to a situation with direct sound, the contribution caused by atmospheric refraction, although small, results in an important increase in the sound pressure level. A similar conclusion could be drawn in the case of sound propagation between adjacent street canyons in an urban environment.[21] The numerical model nevertheless manages to produce sufficiently accurate results.

In both cross sections, the medians of the measured relative sound pressure levels range from −3 to 3 dBA. The relief in combination with the refracting atmosphere compensate for the effect of geometric divergence of the sound wave, ground attenuation and atmospheric absorption.

Note that the calculated relative sound pressure levels also change considerably in the case of a homogeneous atmosphere, because of changes in the magnitude of and the variation in atmospheric attenuation during the periods of the different temperature profile clusters. In the case of temperature profile 4 in cross section 2 (see Fig. 5(d)), the high and constant relative humidity in that period induces almost no atmospheric absorption. The variation in the calculations in that case is caused only by differences in the modeled source spectrum. In the other situations in cross section 2 (see Figs. 5(b) and 5(c)), lower values of relative humidity also occur, leading to more atmospheric absorption. As a result, about 10 dBA difference can be observed between the different clusters when considering the relative sound pressure levels at microphone 24 in the case of a flat and homogeneous atmosphere.

The calculations also allow one to have a closer look at, e.g., frequency spectra. In Fig. 6, the computed sound pressure levels of the 15 most contributing sound paths to the overall sound pressure level at receiver 3 are shown, as a function of frequency. Detailed geometrical information is shown in Figs. 7 and 8: the source-receiver lines plotted on the orthophotos, and the terrain profile between each source point and the receiver. The sound pressure level frequency spectrum for each sound path is clearly different. In Fig. 9, this same information is compacted to total A-weighted sound pressure levels per sound path. The sum of all sound paths (which is the data that are used for the comparison with measurements) is indicated with the horizontal line. The source point closest to the receiver is number 17. Its contribution to the total sound pressure level at the receiver is smaller than, e.g., for sound path 20. On the other hand, an important contribution could have been expected from point 21. As is clear from Fig. 9, this source point is not contrib-
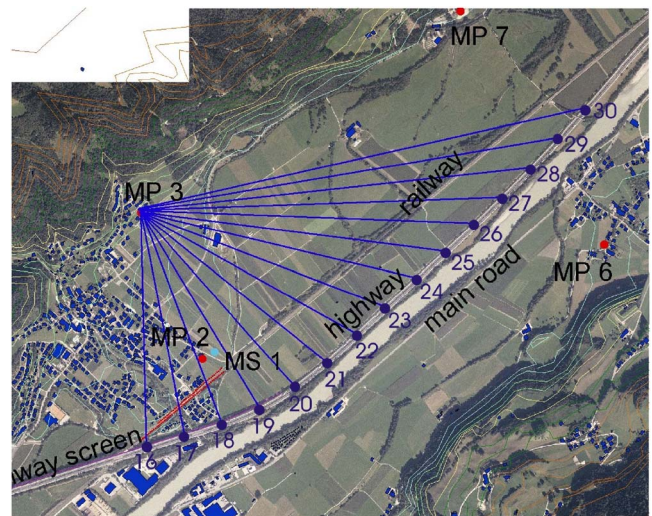
FIG. 7. (Color online) Orthophoto of the 15 source points under consideration for microphone position 3. The source points after refinement are not shown. Height contours are shown on the map.

Van Renterghem *et al.*: Sound propagation in valley-slope configurations
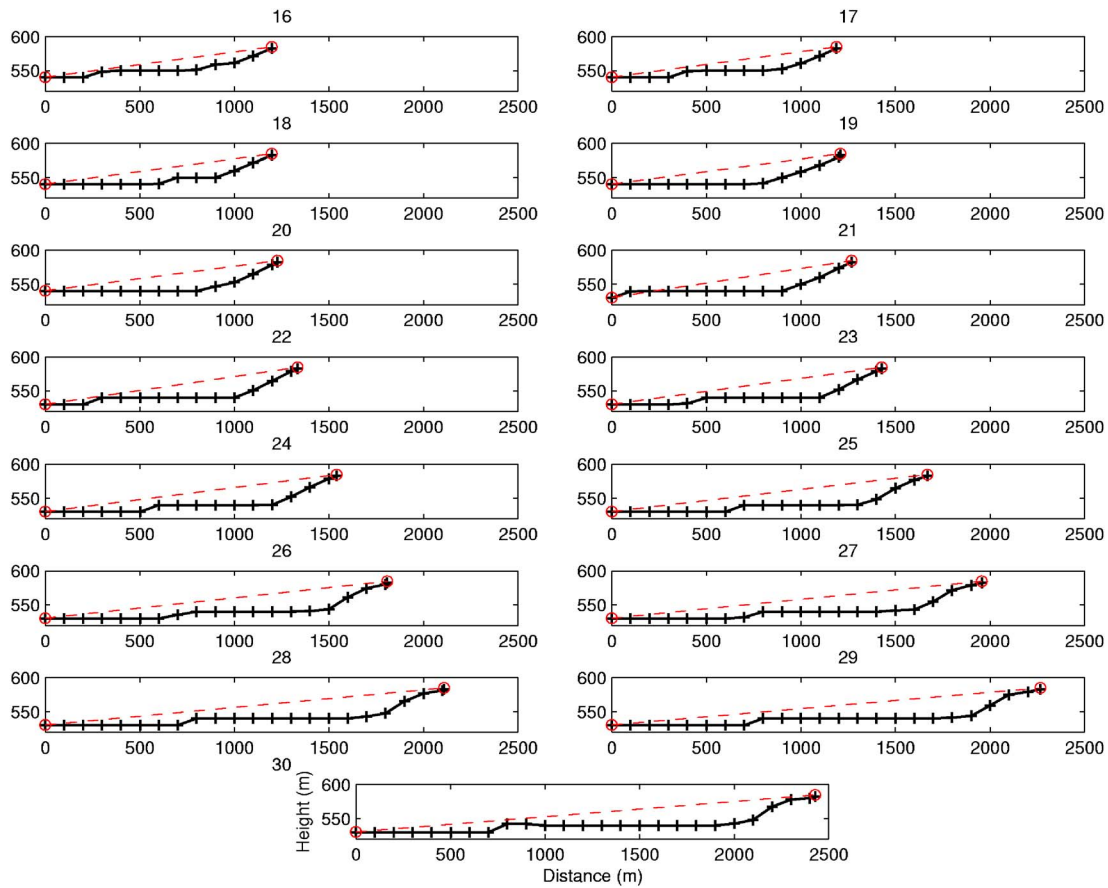
FIG. 8. (Color online) Terrain profiles for the 15 source points under consideration for microphone position 3. The distance towards the source is shown on the horizontal axis, the height is shown on the vertical axis. The source and the receiver (open circles) are connected with a straight, dashed line. The vertical axis is not true to scale. Source points after refinement are not shown.

uting to the sound field at the receiver; the difference in sound pressure level relative to source point 20 is more than 25 dBA. The relief is responsible for this. Starting from source point 21, the road becomes somewhat depressed. As a result, sound is shielded effectively at source point 21: There is no direct view between source and receiver. This analysis

of the source points contributing to microphone position 3 clearly shows that the degree of detail included in our calculations is necessary.

VI. CONCLUSIONS

In this paper, a comparison is made between measured and calculated sound pressure levels in a valley-slope configuration, during windless periods without precipitation. A rotated Green's Function Parabolic Equation (GFrPE) method was used for the numerical calculations. Typical parameters related to GFrPE like the number of actual calculations needed near the transition of successive domains, and the maximum height to be considered on the starting fields, were optimized to increase computational efficiency.

A methodology is presented to calculate sound pressure levels from road traffic. This calculation methodology is related to the HARMONOISE reference model. The road traffic noise source is discretized by a number of point sources, and two-dimensional calculations are performed for each source-receiver vertical plane. The detailed sound propagation calculations include the undulation of the terrain, the presence of obstacles (like noise screens and houses), (ground) impedance discontinuities, and refraction by arbitrary sound speed profiles.

The agreement between numerical calculations and measurements is good; differences are smaller than 3 dBA. The
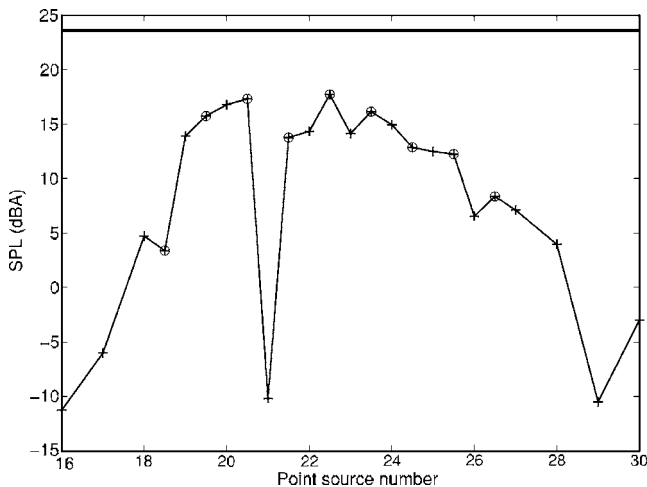
FIG. 9. Total A-weighted sound pressure level resulting from different sound paths, at microphone position 3. A description of the parameters involved is found in the caption of Fig. 6. The source points after refinement are shown with the open circles. The sum of all profiles is indicated with the horizontal line. The magnitudes of the sound pressure levels are arbitrary.
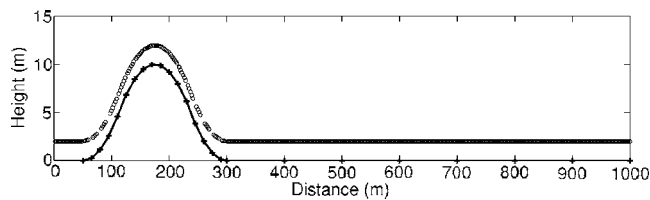
FIG. 10. Terrain profile used for the validation of the GFrPE model. The vertical axis is not true to scale. The open circles indicate the receiver positions.

spread of the sound pressure levels is similar for both the measurements and the calculations. Additional calculations allow one to separate effects from the relief and the refractive state of the atmosphere. The elevation of the terrain is responsible for an increase in the sound pressure level at distant points up to 30 dBA. The temperature profiles observed in this mountainous area are complex and contain upward and downward refracting parts. Their influence ranges from −3 dBA to +10 dBA. The sound pressure levels at the distant points, relative to the reference points, are in most cases near 0 dBA. The relief in combination with the refracting atmosphere compensates the expected decrease in sound pressure level caused by geometric spreading of sound, ground attenuation and atmospheric absorption.

The detailed validation of the GFrPE model for valleys in mountainous areas shows both that the proposed model is accurate enough to predict sound levels up to distances of 1000 m and that further simplification of full wave calculations is not possible. Although computational requirements are strongly reduced by careful tuning of the numerical parameters, the GFrPE method remains too computationally demanding to be used in noise mapping. The method is nevertheless very well suited as a reference model to which an engineering approach—to be used in the noise mapping process—can be validated, also in this particular valley-slope context.

ACKNOWLEDGMENTS

APPENDIX: VALIDATION OF THE GFrPE METHOD

The GFrPE code is validated for the case of the smooth hill presented in Ref. 10. The height of the top is 10 m, and the hill is stretched over a distance of 200 m (see Fig. 10). A reference solution with the General Terrain Parabolic Equation method (GTPE)[22] in this configuration is found in Ref. 10. The effect of this hill on sound propagation is significant, notwithstanding the fact that the height of the top is only
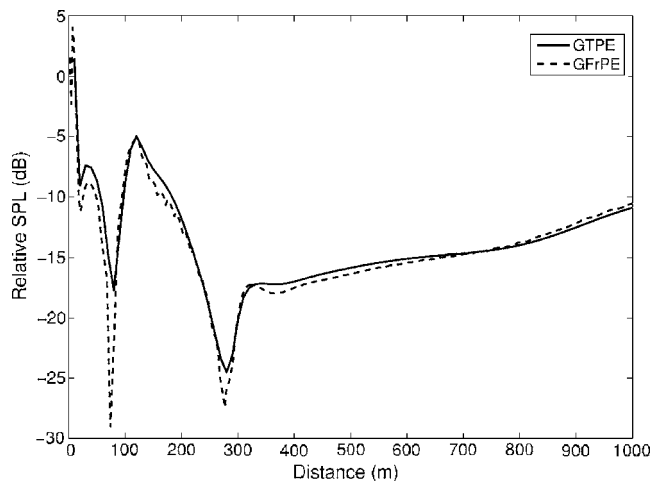
FIG. 11. Comparison between GTPE and GFrPE results in the case of the configuration shown in Fig. 10. The sound pressure level, relative to free field sound propagation, is shown with distance, for a sound frequency of 300 Hz. The one-parameter Delany and Bazley ground impedance model is used with a flow resistivity of 200 kPa s/m^2, in a downward refracting atmosphere.

10 m and that the slope angles are limited. The source is situated 2 m above the ground; a set of receivers is placed 2 m above the ground as well, up to a distance of 1000 m from the source. The sound frequency is 300 Hz. The one-parameter Delany and Bazley ground impedance model[19] is used with a flow resistivity of 200 kPa s/m^2. A logarithmic, sound speed profile $340+\ln(z/0.1+1)$ m/s is used, where z is the height above the ground in meters.

The following parameters are used during the GFrPE calculations. A standard Gaussian starter[5,6] is applied. In vertical direction, ten points per wavelength (λ) are used. The horizontal propagation step is 2λ. Every 15 m, a new flat domain is used to discretize the hill (see Fig. 10). The thickness of the absorbing layer on top of the computational domain is 150λ. Inside this layer, an imaginary term is added to the wave number equal to $iA_t(z-z_t)^2/(z_M-z_t)^2$, where z is the height, z_t is the height where the absorbing layer starts, and z_M is the top of the computational domain. The optimum choice of the constant coefficient A_t depends on frequency and is chosen to be 0.5.[10]

In Fig. 11, a comparison between GFrPE and GTPE is shown. The sound pressure level is expressed relative to free field sound propagation. The agreement between both models is very good. The terrain in the GFrPE calculation is not completely smooth because of the subdivision in flat segments. This causes some small, local distortions of the sound field, especially in the shadow zone of the hill. GFrPE results in a somewhat stronger destructive interference as well.

[1]K. Heutschi, "On the sound propagation in Alpine valleys," *Proceedings of Euronoise 2006*, Tampere, Finland (2006).
[2]C. Whiteman, "Breakup of temperature inversions in deep mountain valleys. Part I. Observations," J. Appl. Meteorol. **21**, 270–289 (1982).
[3]A. Colette, F. Chow, and R. Street, "A numerical study of inversion-layer breakup and the effects of topographic shading in idealized valleys," J. Appl. Meteorol. **42**, 1255–1272 (2003).
[4]D. Heimann and G. Gross, "Coupled simulation of meteorological parameters and sound level in a narrow valley," Appl. Acoust. **56**, 73–100 (1999).

[5]K. Gilbert and X. Di, "A fast Green's function method for one-way sound propagation in the atmosphere," J. Acoust. Soc. Am. **94**, 2343–2352 (1993).

[6]E. Salomons, "Improved Green's function parabolic equation method for atmospheric sound propagation," J. Acoust. Soc. Am. **104**, 100–111 (1998).

[7]P. Blanc-Benon and D. Juve, "Outdoor sound propagation in complex environments: Recent developments in the PE method," *Proceedings of Forum Acusticum 2002*, Sevilla, Spain (2002).

[8]F. Aballéa, Propagation acoustique en milieu extérieur: Application de l'équation parabolique rapide au couplage d'effets météorologiques et de topographies complexes ("Using the fast parabolic equation in complex topographies, including meteorological effects, for sound propagation outdoors"), Ph.D. dissertation, Université de Maine, France, 2004).

[9]E. Salomons, "Diffraction by a screen in downwind sound propagation: A parabolic-equation approach," J. Acoust. Soc. Am. **95**, 3109–3117 (1994).

[10]E. Salomons, *Computational Atmospheric Acoustics* (Kluwer, Dordrecht, The Netherlands, 2001).

[11]T. Van Renterghem, E. Salomons, and D. Botteldooren, "Efficient FDTD-PE model for sound propagation in situations with complex obstacles and wind profiles," Acust. Acta Acust. **91**, 671–679 (2005).

[12]J. Defrance, E. Salomons, I. Noordhoek, D. Heimann, B. Plovsing, G. Watts, H. Jonasson, X. Zhang, E. Premat, I. Schmich, F. Aballea, M. Baulac, and F. de Roo, "Outdoor sound propagation reference model developed in the European Harmonoise project," Acust. Acta Acust. **93**, 213-227 (2007).

[13]ISO 9613-1, "Acoustics–attenuation of sound during propagation outdoors–Part 1," International Organisation for Standardization (Geneva, Switzerland, 1996).

[14]G. Goedecke, R. Wood, H. Auvermann, V. Ostashev, D. Havelock, and C. Ting, "Spectral broadening of sound scattered by advecting atmospheric turbulence," J. Acoust. Soc. Am. **109**, 1923–1934 (2001).

[15]F. Wiener and D. Keast, "Experimental study of the propagation of sound over ground," J. Acoust. Soc. Am. **31**, 724–733 (1959).

[16]P. Chevret, P. Blanc-Benon, and D. Juve, "A numerical model for sound propagation through a turbulent atmosphere near the ground," J. Acoust. Soc. Am. **100**, 3587–3599 (1996).

[17]J. Nicolas, J. Berry, and G. Daigle, "Propagation of sound above a finite layer of snow," J. Acoust. Soc. Am. **77**, 67–73 (1985).

[18]D. Albert and J. Orcutt, "Acoustic pulse propagation above grassland and snow. Comparison of theoretical and experimental waveforms," J. Acoust. Soc. Am. **87**, 93–100 (1990).

[19]M. Delany and E. Bazley, "Acoustic properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).

[20]H. Jonasson and S. Storeheier, *Nord2000*. "New nordic prediction method for road traffic noise," Technical report, SP, SINTEF, 2001. SP Rapport 2001:10.

[21]T. Van Renterghem, E. Salomons, and D. Botteldooren, "Parameter study of sound propagation between city canyons with coupled FDTD-PE model," Appl. Acoust. **67**, 487–510 (2006).

[22]R. Sack and M. West, "A parabolic equation for sound propagation in two dimensions over any smooth terrain profile: The generalised terrain parabolic equation (GTPE)," Appl. Acoust. **45**, 113–129 (1995).

Acoustic monitoring of the tide height and slope-water intrusion at the New Jersey Shelf in winter conditions

Altan Turgut,[a] Marshall Orr, and Bruce Pasewark[b]

Naval Research Laboratory, Acoustics Division, Washington, DC 20375

Waveguide invariant theory is used to describe the frequency shifts of constant acoustic intensity level curves in broadband signal spectrograms measured at the New Jersey Shelf during the winter of 2003. The broadband signals (270–330 Hz) were transmitted from a fixed source and received at three fixed receivers, located at 10, 20, and 30 km range along a cross-shelf propagation track. The constant acoustic intensity level curves of the received signals indicate regular frequency shifts that can be well predicted by the change in water depth observed through tens of tidal cycles. A second pattern of frequency shifts is observed at only 30 km range where significant variability of slope-water intrusion was measured. An excellent agreement between observed frequency shifts of the constant acoustic intensity levels and those predicted by the change in tide height and slope water elevations suggests the capability of long-term acoustic monitoring of tide and slope water intrusions in winter conditions. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2713705]

I. INTRODUCTION

In a shallow-water waveguide, pressure spectra of acoustic signals generated by distant broadband sources often exhibit striation patterns when plotted in time (or range) due to constructive and destructive interference of propagating acoustic modes. For a range-independent waveguide, Chuprov[1] derived a simple formula for the so-called *waveguide invariant* that characterizes the dispersive nature of the waveguide and relates the observed slope of the striations to range and frequency. The range-frequency waveguide invariant, usually designated as beta, has the canonical value of one for an isovelocity waveguide with perfectly reflecting bottom. However, for an arbitrary sound-speed profile, different combinations of modes at different frequencies provide slightly different values of beta. Effects of time- and range-varying waveguide parameters on the striation patterns have also been studied.[2–6] Weston *et al.*[2] observed a change in the frequency-time interference patterns of broadband (4.1–4.5 kHz) signals that could be explained by the dependency of mode parameters on water depth that changed through the tidal cycle. They derived a simple formula for the slope of the interference patterns by using the expression for the modal interaction distance, and showed that measured slopes of interference patterns agree well with the theoretical predictions. Grachev[3] used a unified waveguide invariant theory and showed that the range-frequency invariant beta and several time-frequency waveguide invariant parameters could be used to estimate the slope of constant acoustic intensity levels for more complicated waveguides. Recently, Petkinov and Kuzkin[4] summarized the previously reported theoretical estimations of the frequency shift of constant acoustic intensity, levels produced by tides, internal waves, surface waves, and diurnal oscillations of frontal zone.

Several previous experimental data showed evidence of tidal height effects on the acoustic intensity striation patterns with partial success.[2,7,8] However, acoustic monitoring of a single oceanographic process was never envisioned due to the complexity of environmental conditions in shallow water. Our high-quality experimental data show that some of these oceanographic processes can be isolated and monitored by low-frequency acoustic propagation measurements conducted under winter conditions. In addition to tidal height monitoring, the effects of slope-water intrusion on the measured acoustic intensity striation patterns is demonstrated for the first time by using the same data set collected at the Middle Atlantic Bight. Slope-water intrusion and climatology of shelf break fronts in the Middle Atlantic Bight have been reported based on extensive historical data[9] and most recent oceanographic data from the New England shelf break PRIMER experiment.[10] During the PRIMER experiment, an integrated study of acoustics and oceanography was conducted to understand low-frequency acoustic propagation in the presence of a shelf-break front both in winter and summer conditions.[11] Their oceanographic measurements indicated that dynamics of the shelf break front is strongly influenced by external forcing of Gulf Stream rings and filaments, causing temporal correlation scales of temperature and salinity to be about 1.3 days.[10,12] They also noted a warm near-bottom slope, water layer at the floor of the shelf break front extending shoreward to the 100 m isobath that is also affected by the strong external forcing. Based on numerical simulations, they showed that measured sound-speed fields, representing the dynamics of shelf break fronts, can cause significant acoustic field variability.[11] Our experiment, Relationship between Array Gain and Shelf-break fluid processes (RAGS), was conducted on the outer New Jersey Shelf to further study the interaction of sound with shelf-break fluid processes during winter conditions. The oceanographic data from the RAGS experiment also revealed a warm near-bottom slope-water layer extending shoreward between 80

0001-4966/2007/121(5)/2534/8/$23.00
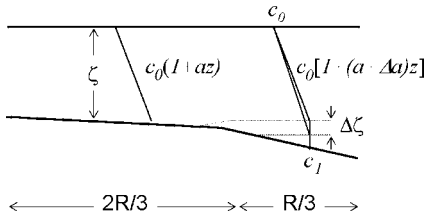
FIG. 1. A simplified RAGS03 experiment acoustic propagation environment representing both ideal waveguide and surface channel propagation conditions.

and 100 m isobaths with about 20 m change in layer thickness within the time scales of one to two days. Observed frequency-time interference patterns of broadband (270–330 Hz) signals showed direct evidence of the effects of these near-bottom slope-water intrusions on low-frequency acoustic propagation.

In the present paper, waveguide invariant theory is used to explain the observed frequency shift of the constant acoustic intensity levels of broadband signals by variations in tide height and thickness of the near-bottom slope water. Acoustic monitoring of the near-bottom slope-water intrusion is demonstrated. In Sec. II, derivations of both range-frequency and time-frequency waveguide invariants are given for a simple waveguide with range- and time-dependent parameters. In Sec. III, numerical simulations are used to investigate the applicability of analytically derived waveguide invariants to the RAGS03 data. In Sec. IV, the RAGS03 New Jersey Shelf experiment is briefly described and observed time-frequency patterns are compared with the theoretical predictions. A summary and discussion are given in Sec. V.

II. THEORY

In this section, we derive various waveguide invariant parameters to describe the observed frequency shift of the constant acoustic intensity levels during the RAGS03 experiment. We first consider the case where the observable is the acoustic intensity from a distant source measured as a function of range and frequency. We derive a specific form of the waveguide invariant that makes explicit its dependence on the propagating acoustic modes. We also consider the case where the source and receiver are fixed in space and the observed acoustic intensity is varied in time due to the time-dependent properties of waveguide. For later use on data from the RAGS03 experiment, we then outline various waveguide invariant parameters for an ideal waveguide and a surface channel (see Fig. 1).

The acoustic intensity, when mapped versus range and frequency, often exhibits a pattern of striations that represent level curves. To analyze the properties of the level curves, let I be the observed intensity as a function of range r from the source and frequency ω. The level curves satisfy

$$\frac{\partial I}{\partial \omega} d\omega + \frac{\partial I}{\partial r} dr = 0. \tag{1}$$

Solving for the slope $d\omega/dr$ of the striations yields

$$\frac{d\omega}{dr} = -\frac{\partial I/\partial r}{\partial I/\partial \omega}. \tag{2}$$

To analyze Eq. (2), it is convenient to use a normal mode representation for the intensity. For source and receiver depths z_s, and z, respectively, the associated pressure p in the far field can be approximated as

$$p = \sum_m (\xi_m r)^{-1/2} \Psi_m(z_s) \Psi_m(z) e^{i(\xi_m + i\alpha_m)r}, \tag{3}$$

where Ψ_m are the mode functions and ξ_m the corresponding horizontal wave numbers. The modal attenuation α_m accounts for bottom loss. Certain unimportant scaling terms have been suppressed. It follows that the intensity, $I = |p|^2$, is

$$I = \sum_{m,n} A_{mn} \exp(i\Delta\xi_{mn} r), \tag{4}$$

where $\Delta\xi_{mn}$ are horizontal wave number differences between modes n and m. The amplitudes A_{mn} can be derived from Eq. (3). Neglecting the weak range and frequency dependence in A_{mn}, the derivatives can be calculated

$$\frac{\partial I}{\partial r} \approx i\omega \sum_{m,n} A_{mn} \frac{\Delta\xi_{mn}}{\omega} \exp(i\Delta\xi_{mn} r), \tag{5a}$$

$$\frac{\partial I}{\partial \omega} \approx ir \sum_{m,n} A_{mn} \frac{\partial \Delta\xi_{mn}}{\partial \omega} \exp(i\Delta\xi_{mn} r), \tag{5b}$$

where $\Delta\xi_{mn}/\omega$ are the phase slowness differences with modal phase velocity $v = (\xi/\omega)^{-1}$ and $\partial\Delta\xi_{mn}/\partial\omega$ are the mode group slowness differences with modal group velocity $u = (\partial\xi/\partial\omega)^{-1}$.

Brekhovskikh and Lysanov[5] give a definition for the waveguide invariant as a derivative of the phase slowness with respect to the group slowness. For a group of closely spaced modes, the derivative of the phase slowness with respect to the group slowness is assumed to be independent of mode order and frequency, and represented by a scalar parameter as[1]

$$\beta = -\frac{\Delta(1/v)}{\Delta(1/u)} = -\frac{\Delta\xi_{mn}/\omega}{\partial\Delta\xi_{mn}/\partial\omega}. \tag{6}$$

Equation (6) can be used in Eq. (5a) to eliminate the term related to the phase slowness difference. Combining Eqs. (2)–(6) yields for the slope of the striations

$$\frac{d\omega}{dr} = \beta \frac{\omega}{r}. \tag{7}$$

Grachev[3] noted that in addition to the above-mentioned range-frequency waveguide invariant, other acoustic field (time-frequency) invariants can be derived for a waveguide with time-dependent parameters, such as waveguide thickness and propagation speed. In this case, similar to Eq. (2), slope of the intensity striations yields

$$\frac{d\omega}{d\eta} = -\frac{\partial I/\partial \eta}{\partial I/\partial \omega}, \tag{8}$$

where η is a time-varying waveguide parameter. Using the normal mode representation of intensity and following a

derivation similar to Eqs. (3)–(6), a time-frequency wave-guide invariant γ can be defined as

$$\gamma = -\frac{\dfrac{\partial \Delta \xi_{mn}}{\partial \eta}/\omega}{\dfrac{\partial \Delta \xi_{mn}}{\partial \omega}/\eta}, \tag{9}$$

from which the slope of the intensity striations can be calculated as

$$\frac{d\omega}{d\eta} = \gamma \frac{\omega}{\eta}. \tag{10}$$

Simultaneous solution of differential equations (7) and (9) for $\Delta \xi_{mn}$ yields[3]

$$\Delta \xi_{mn} = C_{mn} \eta^{\gamma/\beta} \omega^{-1/\beta}, \tag{11}$$

where C_{mn} is a constant that depends only on the mode order.

For an ideal waveguide with an isovelocity sound speed profile, the horizontal component of the wave number is given by[5] $\xi_m = [(\omega/c)^2 - (\pi m/h)^2]^{1/2}$, from which an approximate form of the horizontal wave number difference can be calculated as $\Delta \xi_{mn} \cong \pi^2(m^2 - n^2)(c/2\omega h^2)$. Using the approximate horizontal wave number difference in the above equation yields the values $\beta = 1$ for the range-frequency waveguide invariant, $\gamma_h = -2$ for the time-frequency waveguide invariant of waveguide thickness, and $\gamma_c = 1$ for the time-frequency waveguide invariant of sound speed. In this case, the frequency shift of level curves due to a change in water depth can be calculated from Eq. (10) by setting $\gamma = h$ as

$$\frac{\Delta\omega}{\omega} = -\frac{2\Delta h}{h}. \tag{12}$$

The above-presented formula was first derived by Weston et al.[2] by considering the dependence of modal interaction distance on the water depth. He then used the above-noted formula to explain the observed slopes of the interference patterns during a half tidal cycle.

For a surface channel with a linear variation in the square index of refraction, $n^2(z) = c_0^2/c^2(z) = 1 - az$, the horizontal wave number differences in Eq. (11) is given as

$$\Delta \xi_{mn} \approx \tfrac{1}{2}(y_m - y_n)a^{2/3}c_0^{1/3}\omega^{-1/3}, \tag{13}$$

where y_m and y_n are the roots of the Airy function. Equation (13) yields the values $\beta = -3$ for the range-frequency waveguide invariant, $\gamma_a = -2$ for the time-frequency waveguide invariant of the sound-speed gradient, and $\gamma_{c_0} = 1$ for the time-frequency waveguide invariant of the sound speed. In the case where $a \ll 1$, assuming a linear variation in the index of refraction, the sound speed profile can be approximated as $c(z) \approx c_0(1 + az/2)$.

Calculations of horizontal wave number differences can be found in Kibblewhite and Denham[13] for an upward refracting surface channel or in Kuzkin[14] for a downward refracting near-bottom channel in the form

$$\Delta \xi_{mn} \approx \frac{(3\pi)^{2/3}}{2}[(m - 3/4)^{2/3} - (n - 3/4)^{2/3}]a^{2/3}c_0^{1/3}\omega^{-1/3}. \tag{14}$$

Equation (14) also yields the values $\beta = -3$ for the range-frequency waveguide invariant, $\gamma_a = -2$ for the time-frequency waveguide invariant of the sound-speed gradient, and $\gamma_{c_0} = 1$ for the time-frequency waveguide invariant of the sound speed. Let us consider a surface channel with thickness ζ, surface sound speed c_0, and sound speed gradient a overlaying a bottom layer with a constant sound speed c_1 (see Fig. 1 for the region $2R/3 < r < R$ where R is the distance between the fixed source and receiver). We assume that both refracting modes (trapped in the surface channel) and reflecting modes (bottom/surface) exist in the waveguide. For trapped modes, using sound speed gradient $a = (c_1 - c_0)/c_0\zeta$ and its variation $\Delta a = -\Delta\zeta(c_1 - c_0)/c_0\zeta^2$, the time-frequency waveguide invariant for surface duct thickness can be related to the time-frequency waveguide invariant for sound-speed gradient as $\gamma_s = -\gamma_a$. In this case, the frequency shift due to a change in surface duct thickness can be calculated from Eq. (14) as

$$\frac{\Delta\omega}{\omega} = \frac{2\Delta a}{2} = -\frac{2\Delta\zeta}{\zeta}. \tag{15}$$

For the (bottom/surface) reflecting modes, Eq. (12) can also be used to calculate the frequency shift due to change in water depth. Note that in the case of gradual range dependency of the waveguide as shown in Fig. 1. Equation (9) can be extended as[6]

$$\gamma(r) = -\frac{\displaystyle\int_0^r \frac{\partial \Delta \xi_{mn}}{\partial \eta}dr/\omega}{\displaystyle\int_0^r \frac{\partial \Delta \xi_{mn}}{\partial \omega}dr/\eta}. \tag{16}$$

For the (bottom/surface) reflecting modes at ranges $0 < r < 2R/3$, the frequency shift can be calculated from Eq. (16) as[6]

$$\frac{\Delta\omega}{\omega} = -\frac{2\Delta h}{h_{\text{eff}}}, \tag{17}$$

where $h_{\text{eff}} = (h_s h_r)^{1/2}$ is the effective water depth,[15] h_s is the water depth at the source, and h_r is the water depth at the receiver. Due to slope-water intrusion, if the surface waveguide thickness ζ is temporally varied only at ranges $2R/3 < r < R$ as shown in Fig. 1, the frequency shift due to a change in surface duct thickness can be calculated from Eq. (16) as

$$\frac{\Delta\omega}{\omega} = -\frac{2\Delta\zeta}{3\zeta}. \tag{18}$$

III. NUMERICAL SIMULATIONS

In this section, acoustic intensity patterns are calculated numerically for the propagation environment depicted in Fig. 1. Waveguide invariants derived in Sec. II are compared with
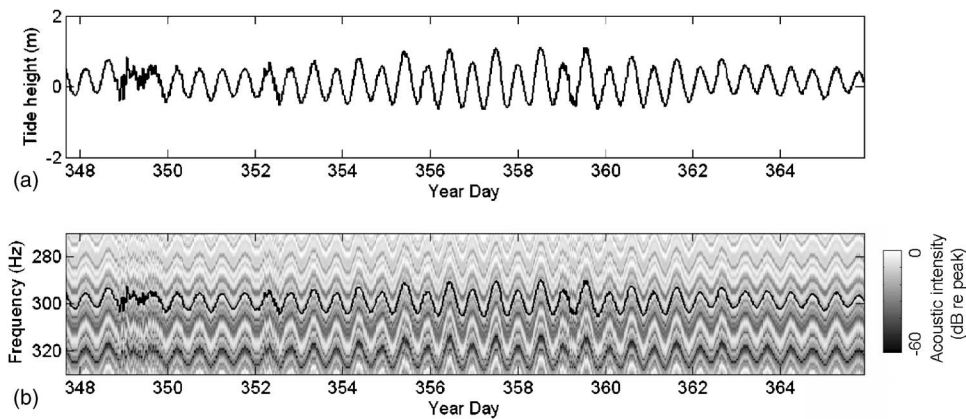
FIG. 2. (a) Measured tide height at 20 km range and (b) simulated acoustic intensity levels using the measured tide during the period from YD348 to YD365. The frequency shifts predicted by using Eq. (17) is also shown (solid line).

those calculated from the simulated data. Numerical simulations are performed using a broadband adiabatic normal-mode propagation model that uses the modal eigenvalues and eigenfunctions generated by the numerical code FEMODE.[16] First, frequency shifts of 270–330 Hz signals are simulated at 20 km range by changing the water depth according to the measured tidal information during the period from YD347 to YD365 (see Fig. 2). The source is located 19 m above the bottom in 65 m water depth. The water depth is linearly increased with range so that the water depth at the receiver (20 km range) is 89 m. Sound speed profiles in the water column are described in Fig. 1 within the region $0 < r < 2R/3$ where $c_0 = 1484$ m/s at the surface and $c_1 = 1502$ at 60 m depth. Throughout the calculations, the same bottom sound-speed profiles and attenuation values were used. The bottom sound speed is described as $c_b = 1580$ m/s at the seafloor and 1680 m/s at 20 mbsf (meters-below-seafloor) and 1890 m/s at 50 mbsf and below. A value of 0.1 dB per wavelength was used for the bottom attenuation coefficient.

Figure 2(a) shows measured tidal variation of the water depth at 20 km propagation range. Figure 2(b) shows calculated acoustic intensity levels at a receiver located at 20 km range and 20 m depth for the period from YD347 to YD365. The frequency shifts due to measured tidal height variations are also calculated for the frequency $f = 300$ Hz by using Eq. (17) and plotted in Fig. 2(b) (thick solid line). Note the excellent agreement between the simulated frequency shifts and those calculated by using Eq. (17). According to Eq. (17), if the effective water depth is $h_{eff} = (h_s h_r)^{1/2} = 76$ m, and typical tidal height variation is $\Delta h = 0.75$ m, a frequency shift

of $\Delta f = 6$ Hz is expected at the frequency $f = 300$ Hz. This frequency shift can be easily measured with high resolution as is shown in Fig. 2(b) confirming the feasibility of acoustic monitoring of tide height on the continental shelves.

As a second case, dependency of the frequency shifts on the near-bottom (slope-water) layer thickness variations is analyzed by numerical simulations. The same environmental parameters were used as in the previous case, except that the thickness of the surface channel is changed in time according to values extracted from measured temperature profiles at 20 km range [see Fig. 3(a)]. Figure 3(b) shows the calculated acoustic intensity levels at a receiver located at 30 km range and 25 m depth for the period from YD347 to YD365. The frequency shifts due to measured changes in the thickness of the surface layer (or near-bottom slope water layer) are also calculated for the frequency $f = 300$ Hz by using Eq. (18) and plotted in Fig. 3(a) (thick solid line). There is an excellent agreement between the simulated frequency shifts and those calculated by using Eq. (18). According to Eq. (18), if the mean thickness of the surface layer is $\zeta = 85$ m and the thickness variation is $\Delta \zeta = 20$ m, a frequency shift of $\Delta f = 23.5$ Hz is expected at the frequency $f = 300$ Hz. This frequency shift is easily measured with high resolution, as is shown in Fig. 3(b), confirming the feasibility of acoustic monitoring of the change in the thickness of near-bottom slope water. In Fig. 3(a), the change in surface channel thickness exceeds 25 m during several time periods (e.g., at YD361). One would need to use acoustic signals with bandwidths larger than 60 Hz to monitor these extreme cases.

Next, the depth variation of the range-frequency waveguide invariant beta is analyzed to isolate the effects of re-
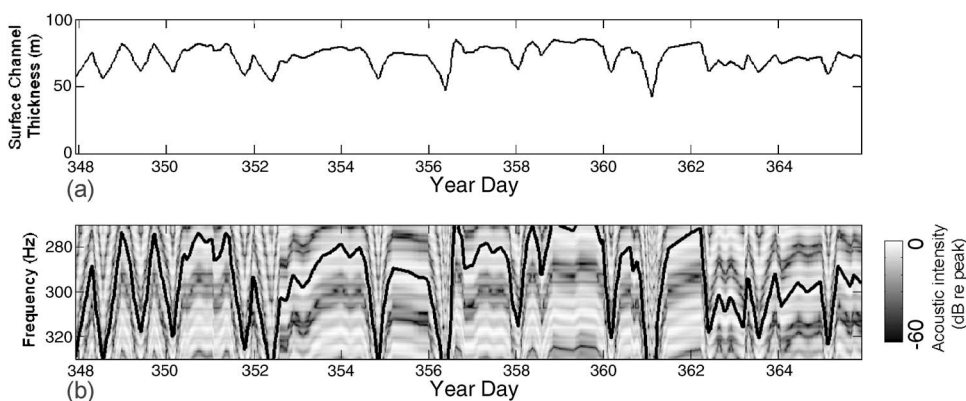
FIG. 3. (a) Surface channel thickness extracted from measured temperature profiles and (b) simulated acoustic intensity levels using the measured tide during the period from YD348 to YD365. The frequency shifts predicted by using Eq. (18) are also shown (solid line).
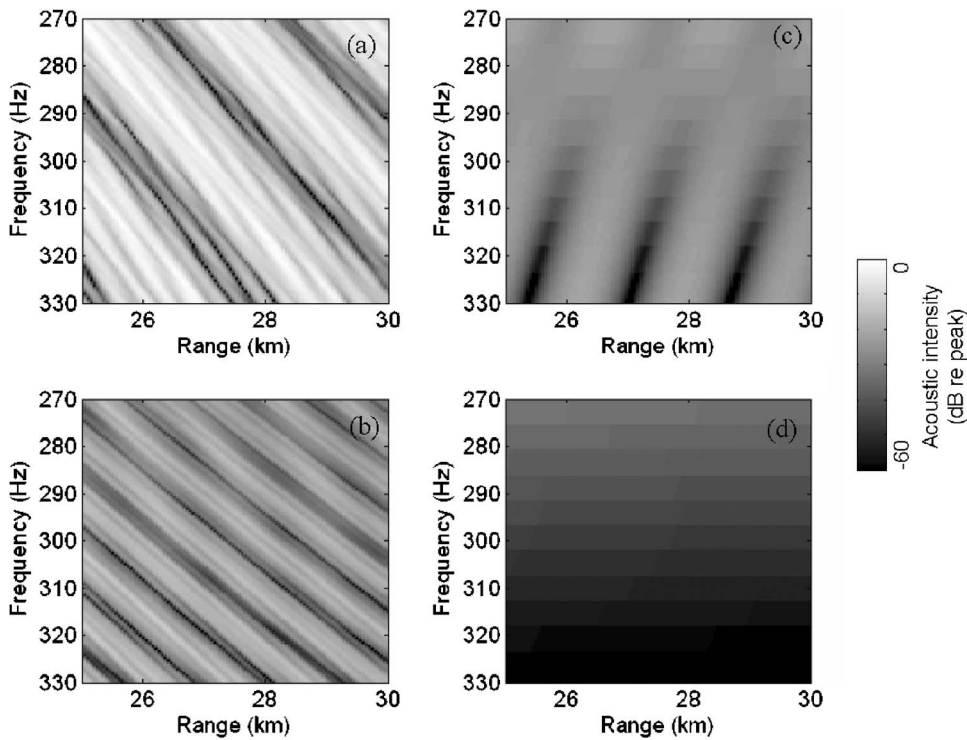
FIG. 4. Simulated acoustic intensity levels for a source moving from 25 to 30 km range using (a), (b) all propagating modes and (b), (d) the first two modes only. (a), (c) Receiver in the surface channel (at 25 m). (b), (d) Receiver below the surface channel (at 95 m).

flecting and refracting modes in a waveguide having both an upward refracting surface channel and an isovelocity near-bottom layer (the region $2R/3 < r < R$ in Fig. 1). Figure 4(a) shows the calculated acoustic intensity levels for a receiver at 25 m depth as the source moved from 25 to 30 km range. The complexity of the acoustic intensity striation patterns indicates that interference of more than one group of acoustic modes might be involved. Figure 4(b) shows the calculated spectral levels for a receiver at 95 m depth. In this case, the acoustic intensity striation patterns are much more regular and the interference of only one group of modes is dominant. For this case, using $r = 27.5$ km, $\Delta r = 5$ km, $f = 300$ Hz, and $\Delta f = 55$ Hz, Eq. (7) yields a value of $\beta \approx 1$ for the range-frequency waveguide parameter. Figures 4(c) and 4(d) show the calculated spectral levels at 25 and 95 m depths that use the first two modes only. In Fig. 4(d), the contribution of the first two modes to the total received levels [see Fig. 4(b)] is minor for the receiver at 95 m depth. Nonexistence of the striation patterns indicates that the relative contribution from one of the first two modes is much weaker than the other. In Fig. 4(c) the slopes of the acoustic intensity striations are reversed for the receiver at 25 m depth with respect to the total field case [Fig. 4(a)] indicating the existence of a group of low-order acoustic modes that are trapped in the surface channel. For this case, using $r = 29$ km, $\Delta r = 1.9$ km, $f = 300$ Hz, and $\Delta f = -60$ Hz, Eq. (7) yields a value of $\beta \approx -3$ for the range-frequency waveguide parameter.

IV. EXPERIMENTAL RESULTS

The Relationship between Array Gain and Shelf-break fluid processes (RAGS) experiment was conducted on the New Jersey Shelf between 12 December 2003 and 4 January 2004. The primary objective was to study the interaction of sound with shelf-break fluid processes in the ocean wave-guide. The collection of high-quality acoustic and oceanographic data provided new analysis and interesting findings in additional research areas.[17,18] In this paper, a unique observation of the frequency shifts of the constant acoustic intensity level curves and their relation to tidal height and slope-water intrusion are explained by using the waveguide invariant theory. The experiment featured moored sources transmitting 16, 60, and 100 Hz bandwidth acoustic signals at respective center frequencies of 224, 300, and 500 Hz. Three vertical receiving arrays, each having 32 elements, were located at ranges 10, 20, and 30 km from the sources (see Fig. 5). Only the 300 Hz signals with 60 Hz bandwidth are considered in this paper. Figures 6(a)–6(c) show the acoustic intensity levels measured at 30 m depth at 10, 20, and 30 km range, respectively. The regular sinusoidal inter-
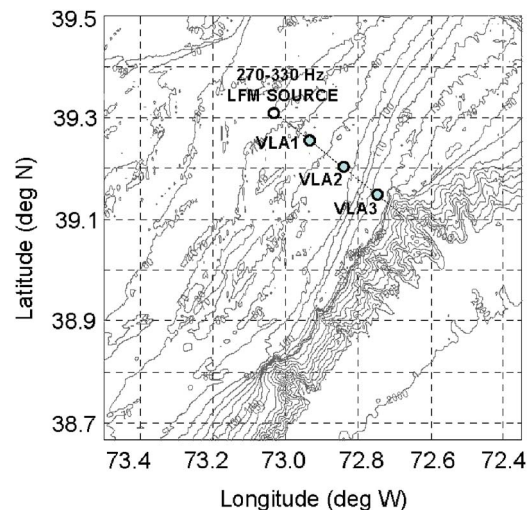
FIG. 5. (Color online) RAGS03 experimental area and source/receiver configuration.
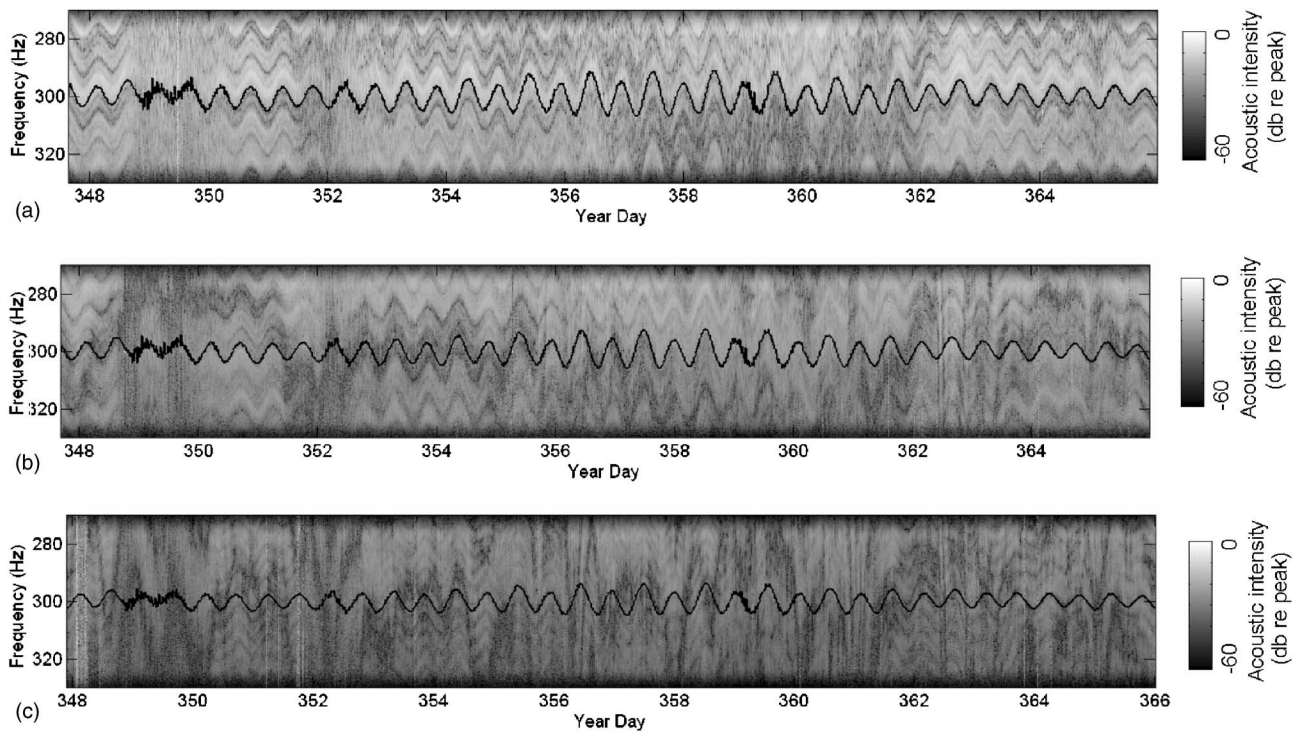
FIG. 6. Measured acoustic intensity levels at (a) 10 km, (b) 20 km, and (c) 30 km range. Corresponding frequency shifts predicted by using Eq. (17) are also shown (solid lines).

ference patterns can be observed throughout the experimental period from YD347 to YD365. The interference patterns slightly deteriorate due to an increase in the range-dependency of waveguide for larger receiver distances. Also, as the effective depth increases with range, the amplitude of the sinusoidal patterns becomes smaller for more distant receivers. The frequency shifts due to the measured tidal height variations [see Fig. 2(b)] are calculated by using the corresponding effective depths 67, 76, and 91 m in Eq. (17) and are also plotted in Figs. 6(a)–6(c) (solid line), respectively. Figures 6(a) and 6(b) show an excellent agreement between the measured acoustic intensity level curves and those calculated by using Eq. (17). Conversely, this also indicates that one can monitor tidal height variation from the measured acoustic intensity level curves. Figure 6(c) shows that acoustic intensity striation patterns are much more complex at 30 km range and the direct effects of tidal height variations can be seen only during certain periods (e.g., at YD350, YD353, and YD361). This indicates significant contributions from the variability of waveguide parameters other than that of tidal height. Figure 6(c) also depicts a second pattern of much larger frequency shifts that can be related to the near-bottom slope water intrusion. Details of the RAGS03 oceanographic measurements are described next to demonstrate the effects of slope water intrusion on the measured frequency shifts of acoustic intensity level curves.

The hydrographic measurements during the RAGS03 experiment include thermistor chains located at each VLA and several days of towed CTD surveys along the acoustic propagation tracks. Figures 7(a) and 7(b) show two towed CTD casts along the acoustic propagation track taken on YD345 and YD347. The near-bottom slope water intrusion

extends shoreward from the 100 m isobath to the 80 m isobath within two days. Also the 1500 m/s isospeed contour is elevated about 30 m, indicating an increase of near-bottom slope water layer thickness. In Fig. 7(c), representative sound speed profiles, plotted at 22 km range, depict an upward refracting surface channel and an isospeed near-bottom layer. Unfortunately, towed CTD casts were limited to only a few days so that monitoring of slope water intrusion for the entire experiment duration was not possible by towed CTD surveys. However, thermistor chains at each VLA provided point measurements of temperature fields that were used to predict the extent of the slope water intrusion from YD437 to YD365.

Figure 8(a) shows the measured temperature field at VLA2 (20 km range) depicting an increase in the warm near-bottom layer thickness, mainly due to the external forcing of the shelf break front by Gulf Stream eddies and filaments. Figure 8(b) shows the acoustic intensity levels measured at VLA3 (30 km range) and 30 m depth. The solid line in Fig. 8(a) is manually traced at a maximum temperature gradient to depict the magnitude of change in the near-bottom layer thickness. The frequency shifts due to the measured near-bottom layer thickness are calculated by using Eq. (18) and are also plotted in Fig. 8(b). For the duration of the entire experiment, an excellent agreement has been observed between the measured acoustic intensity level curves and those estimated by using Eq. (18). This reveals the long-term monitoring capability of the near-bottom layer thickness from the acoustic intensity level curves measured during winter conditions.
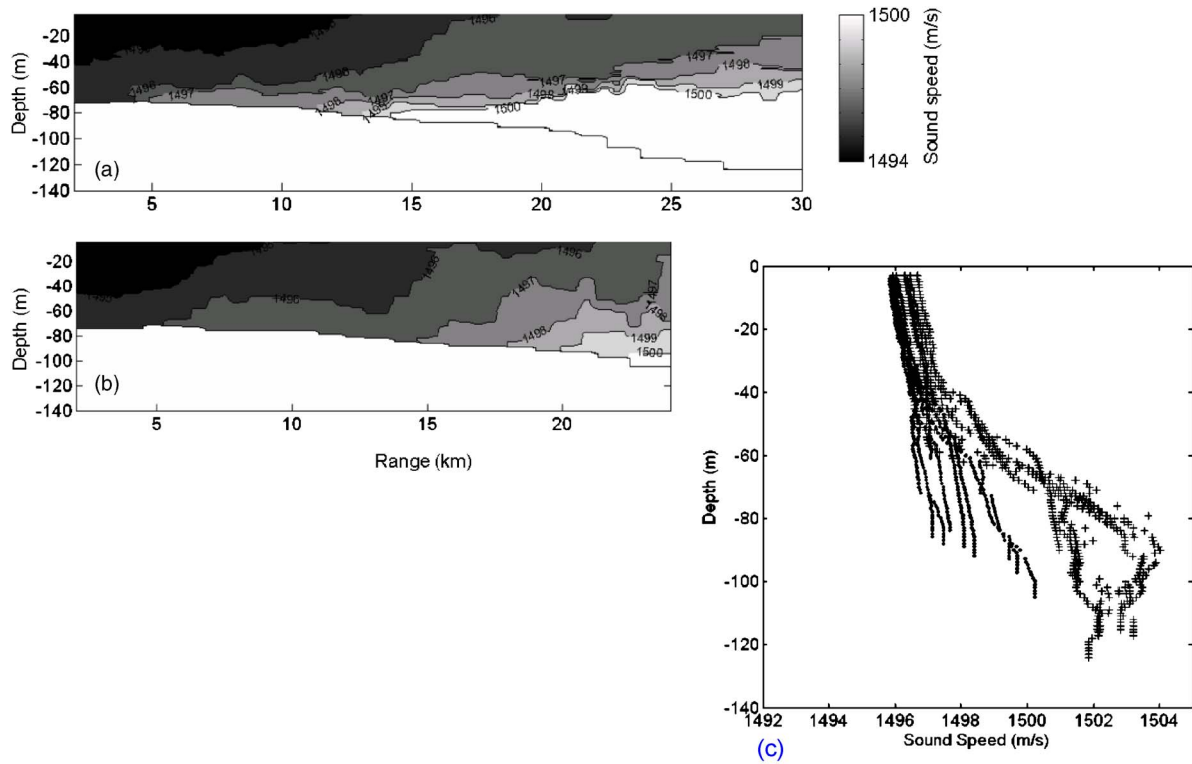
FIG. 7. (Color online) Sound speed fields along the acoustic propagation track extracted from SeaSoar casts taken on (a) YD345, (b) YD347, and (c) representative sound speed profiles near 22 km range.

V. SUMMARY

The relationship between changes in tide height and slopes of acoustic intensity level curves has been reported previously[2,12] with very limited experimental data. In this paper, we have demonstrated the robustness and reliability of this relationship by the high-quality measurements of the tide height and acoustic intensity level curves for much longer durations (tens of tidal cycles) during winter conditions. Also, we have shown the direct effects of slope-water intru-

sion on the frequency shifts of acoustic intensity level curves. The theoretical predictions, given in Sec. II, provide an accurate estimation of the frequency variation of acoustic intensity level curves from the variations of tidal height and near-bottom layer thickness measured in the Middle Atlantic Bight. The excellent agreement between measured frequency shifts of intensity level curves and those by predicted from measured tide height and near-bottom layer thickness indicates the feasibility of long-term acoustic monitoring of tide
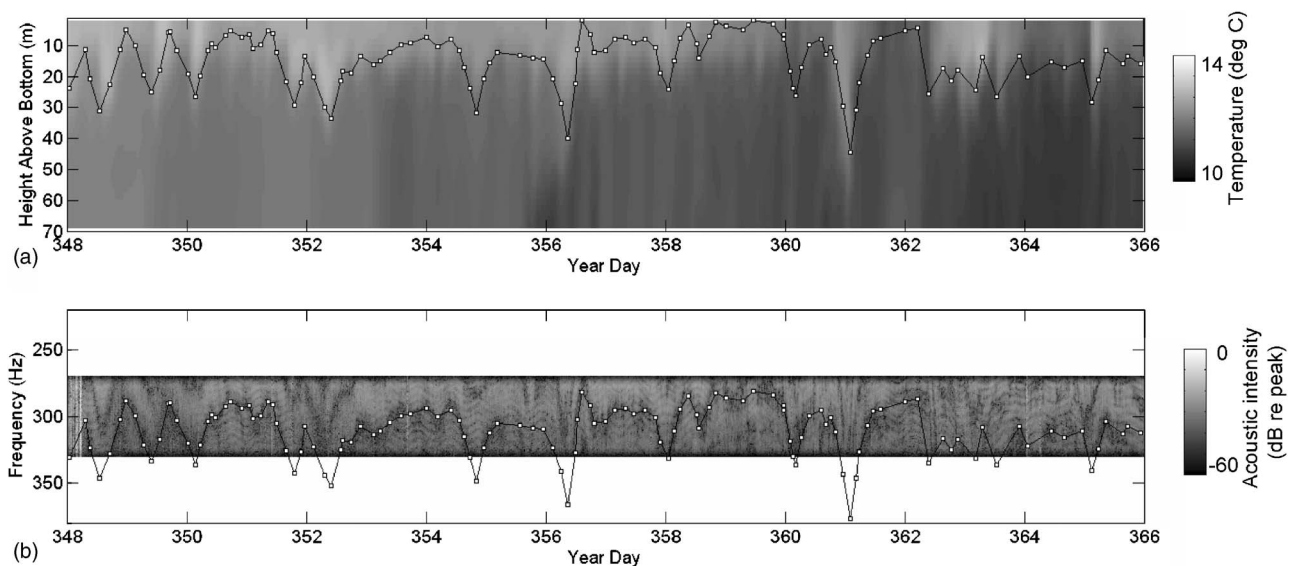
FIG. 8. (a) Measured temperature field at 20 km range and manually traced maximum temperature gradient (solid line), (b) acoustic intensity levels measured at 30 km range and 30 m depth. The frequency shifts predicted by using Eq. (18) are also shown (solid line).

height and near-bottom slope water intrusion in the Middle Atlantic Bight under winter conditions. A natural extension of our results would be the estimation of the effective depth of the acoustic propagation track from the tidal height data obtained from oceanographic measurements or tide prediction models. The effective depth is one of the inverted parameters in the geoacoustic inversion methods commonly used in shallow water.[15] Another potential application of our results is that a smoothly varying two-dimensional (2D) bathymetry can be mapped and monitored by means of an acoustic tomography experiment by collecting broadband acoustic propagation data. Analysis of the 2D mapping capability of the bathymetry and slope-water intrusion is beyond the scope of this paper and results of a low-cost tomography experiment are planned to be reported in the near feature.

ACKNOWLEDGMENTS

[1] S. D. Chuprov, "Interference structure of sound field in the layered ocean," in *Ocean Acoustics. Modern State*, edited by L. M. Brekhovskikh and I. B. Andreeva (Nauka, Moscow, 1982), pp. 71–91.

[2] D. E. Weston, D. Smith, and G. Wearden, "Experiments on time-frequency interference patterns in shallow-water acoustic transmission," J. Sound Vib. **10**, 424–429 (1969).

[3] G. A. Grachev, "Theory of acoustic field invariants in layered waveguides," Acoust. Phys. **39**, 33–35 (1993).

[4] V. G. Petnikov and V. M. Kuz'kin, "Shallow water variability and its manifestation in the interference pattern of sound fields," in *Ocean Acoustic Interference Phenomena and Signal Processing*, edited by W. A. Kuperman and G. L. D'Spain (AIP Conference Proceedings, Melville, NY, 2002), pp. 207–217.

[5] L. M. Brekhovskikh and Y. P. Lysanov, *Fundemantals of Ocean Acoustics*, 2nd ed. (Springer, New York, 1991).

[6] G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1991).

[7] V. M. Kuz'kin, A. V. Ogurtsov, and V. G. Petkinov, "The effect of hydrodynamic variability on frequency shifts of the interference pattern of a sound field in a shallow sea," Acoust. Phys. **44**, 94–100 (1998).

[8] L. F. Bonder, L. K. Bugeva, and A. N. Rutenko, "Effect of tide on sound propagation in the shelf zone of the Sea of Japan," Acoust. Phys. **46**, 534–543 (2000).

[9] C. A. Linder and G. Gawarkiewicz, "A climatology of the shelfbreak front in the Middle Atlantic Bight," J. Geophys. Res. **103**, 18,405–18,423 (1998).

[10] G. Gawarkiewicz, K. H. Brink, F. Bahr, R. C. Beardsley, M. Caruso, J. F. Lynch, and C.-S. Chui, "A large-amplitude meander of the shelfbreak front during summer south of New England: Observations from the shelfbreak PRIMER experiment," J. Geophys. Res. **109**, 1468–1484 (2004).

[11] J. F. Lynch, A. E. Newhall, B. Sperry, G. Gawarkiewicz, A. Frederics, P. Tyack, C.-S. Chiu, and P. Abbot, "Spatial and temporal variations in acoustic propagation characteristics at the New England shelfbreak front," IEEE J. Ocean. Eng. **28**, 129–150 (2003).

[12] R. S. Pickart, D. J. Torres, T. K. McKee, M. J. Caruso, and J. E. Przystup, "Diagnosing a meander of the shelf break current in the Middle Atlantic Bight," J. Geophys. Res. **104**, 3121–3132 (1999).

[13] A. C. Kibblewhite and R. N. Denham, "Experiment on propagation in surface sound channels," J. Acoust. Soc. Am. **38**, 63–71 (1965).

[14] V. M. Kuz'kin, "Frequency shifts of the sound field interference pattern in a shallow sea," Acoust. Phys. **45**, 224–229 (1999).

[15] C. H. Harrison and M. Siderius, "Effective parameters for matched field geoacoustic inversion in range-dependent environments," IEEE J. Ocean. Eng. **28**, 432–445 (2003).

[16] M. D. Collins, "FEPE user's guide," Naval Ocean Research and Development Activity, Stennis Space Center, MS, NORDA Technical Note No. 365, 1988.

[17] A. Turgut, "Geoacoustic inversion by using broadband ship noise recorded on the New Jersey Shelf," J. Acoust. Soc. Am. **118**, 1857 (2005).

[18] A. Turgut, "Acoustic monitoring of finback whale movements on the New Jersey Shelf," J. Acoust. Soc. Am. **120**, 3266 (2006).

Statistical description of chaotic rays in a deep water acoustic waveguide

A. L. Virovlyansky,[a] A. Yu. Kazarova, and L. Ya. Lyubavin
Institute of Applied Physics, Russian Academy of Science, 46 Ul'yanov Street, 603950 Nizhny Novgorod, Russia

This paper analyzes the chaotic ray dynamics at multimegater ranges in a deep water environment with internal-wave-induced fluctuations of the sound speed. The behavior of acoustic ray paths is investigated using the Hamiltonian formalism expressed in terms of action-angle variables. It is shown that the range dependence of the action variable of chaotic ray can be approximated by a random Wiener process. On the basis of this result an approximate statistical description of the chaotic ray structure is derived. Distributions of coordinates, momenta (grazing angles), and actions of sound rays are evaluated. This statistical approach is used for studying ray travel times, that is, arrival times of sound pulses coming to the receiver through different ray paths. The spread of travel times for a bundle of rays with close starting parameters and the influence of sound speed fluctuations on the timefront representing ray arrivals in the time-depth plane are examined. Estimates for the widening and bias of the timefront segment caused by the fluctuations are obtained. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717429]

I. INTRODUCTION

In recent years it has been demonstrated that the ray-based description of the sound field in a deep water acoustic waveguide can properly predict many important features of the arrival pattern at megameter ranges. Numerical results obtained in the scope of the geometrical optics approximation are consistent with both parabolic-equation-based simulations and field experiments.[1–3] Simulations show that even at distances $O(1000 \text{ km})$ effects of ray scattering may dominate diffractive effects.[4] These results stimulate the developing of ray-based approaches for the analysis of long-range sound propagation.

In what follows we shall study the ray structure of acoustic field in a simple but realistic two-dimensional environmental model. The sound speed field $c(r,z)$ (r is range and z is depth) in this model is presented in the form

$$c(r,z) = c_0(z) + \delta c(r,z), \tag{1}$$

where $c_0(z)$ is a smooth (unperturbed) background profile and $\delta c(r,z)$ is a weak range-dependent perturbation caused by random internal waves.[5] Due to this perturbation the ray dynamics at long ranges becomes stochastic and therefore it should be treated by statistical methods. Traditional approaches for describing stochastic rays are associated with ideas of the study of wave propagation in random media (WPRM).[5,6]

In the present paper we consider an alternative approach for the analysis of the irregular ray dynamics. It was suggested by the optical-mechanical analogy and appeared in underwater acoustics in the beginning of 1990s.[7–9] The point is that the motion of a ray path is governed by practically the same Hamilton equations as the motion of a nonlinear oscillator driven by a nonstationary force. In mechanics it is well known that the oscillator typically exhibits chaotic behavior.[10] Its analysis is a classical problem in the theory of dynamical chaos.[11–13] Numerical simulations demonstrate that ray paths in range-dependent underwater waveguides behave in a similar way.[4,14] Chaotic rays are very unstable: Trajectories with close starting parameters diverge exponentially with range. The difference in their vertical coordinates, Δz, grows (on average) with range r as[3,7,15,16]

$$\Delta z \sim e^{\beta r}, \tag{2}$$

where β is the Lyapunov exponent. In an environmental model with a realistic internal-wave-induced perturbation $\delta c(r,z)$ the Lyapunov exponents β are on the order of 0.01 km^{-1}.[3,4] At ranges $O(1000 \text{ km})$ the ray chaos is well developed and this phenomenon should be taken into account when analyzing the statistics of the wave field.[4,14,16]

The theory of WPRM and the theory of ray chaos investigate the ray structure from different and complementary viewpoints. Let us formulate the difference in statements of problems in these two approaches.

(i)   The description of WPRM is based on the notion of statistical ensemble. The latter consists of infinitely many realizations of the waveguide specified by different $\delta c(r,z)$. Statistical characteristics of a ray with given starting parameters z_0 and χ_0—initial depth and grazing angle, respectively—are determined by averaging over rays with the same initial parameters in all realizations.

(ii)   The theory of ray chaos deals with a *deterministic* medium. In our case the latter is specified by a *single* realization of random perturbation. At ranges $r \gg \beta^{-1}$ initially close ray paths become practically independent and the averaging over their starting parameters can be considered as the statistical averaging.

[a]Electronic mail: viro@hydro.appl.sci-nnov.ru

An approximate analytical description of chaotic ray structure in a deep water environment was derived in Refs. 17–19. This was done using the Hamiltonian formalism expressed in terms of the action-angle variables (I, θ). These variables are a convenient tool for studying oscillations of particles (in mechanics) and rays (in wave theory).[7,20] The action variable I determines both ray amplitude and cycle length while the angle variable θ (it should not be confused with the ray grazing angle) determines the position of a current ray point within the cycle. The angle variable θ may be interpreted as a phase of an oscillating ray path.

The statistical description of rays in a waveguide with a perturbation caused by random internal waves is greatly simplified due to the fact that θ rapidly randomizes and already at comparatively short ranges (hundreds of kilometers) for most rays it becomes more or less uniformly distributed on $(0, 2\pi)$. Then the problem reduces to investigation of a slow diffusion of action I described by the Fokker-Planck equation. A similar approach is commonly applied in the studies of dynamical chaos.[11] In Refs. 17–19 it was shown that irregular range variations of action in a realistic environmental model may be approximated by a random Wiener process representing the simplest model of diffusion.[21]

In the following we discuss the statistical approach derived in Refs. 17–19. We demonstrate how it can be used for analyzing distributions of coordinates, momenta (grazing angles), actions, and travel times of chaotic rays. Most analytical estimates obtained in the scope of our approach are valid only for steep rays. In this paper it is shown that there exists a simple numerical procedure for treating flat rays.

The paper is organized as follows. In Sec. II, a brief description of the Hamiltonian formalism in terms of the momentum-position and action-angle canonical variables is given. Our statistical approach for description of chaotic rays derived in Refs. 17–19 is formulated in Sec. III. In Sec. IV this approach is applied to the evaluation of the probability density functions of different ray parameters. Section V is concerned with statistics of ray travel times. Our attention is focused on the analytic description of spread of ray travel times under conditions of ray chaos. In Sec. VI the results of this work are summarized.

II. HAMILTONIAN FORMALISM

A. Momentum-position variables

Consider a deep water waveguide with the sound speed field determined by Eq. (1). It is assumed that the z axis is directed upward and the sea surface is located at $z=0$. In the scope of the Hamiltonian formalism a ray trajectory at range r is characterized by its current coordinate z and momentum p.[3,4,7,14] The latter is related to the ray grazing angle χ through $p = \tan \chi$. The ray dynamics is determined by the Hamiltonian

$$H = \frac{p^2}{2} + U(r,z),\tag{3}$$

where

$$U(r,z) = \frac{1}{2}\left(1 - \frac{c_r^2}{c^2(r,z)}\right)\tag{4}$$

is an analog to potential in classical mechanics, and c_r is a reference sound speed. Let us present the "potential" U as a sum $U(r,z) = U_0(z) + V(r,z)$, where $U_0(z)$ is an unperturbed constituent defined by Eq. (4) with $c(r,z)$ replaced by $c_0(z)$ and

$$V(r,z) = U(r,z) - U_0(z) \simeq \delta c(r,z)/c_r\tag{5}$$

is a small perturbation. The last equality implies that $|c(r,z) - c_r| \ll c_r$. The Hamilton (ray) equations have the form

$$\frac{dp}{dr} = -\frac{dU_0}{dz} - \frac{\partial V}{\partial z}, \quad \frac{dz}{dr} = p.\tag{6}$$

The ray travel time t is given by the integral

$$c_r t = \int_\Gamma (p\,dz - H\,dr)\tag{7}$$

over a ray path Γ. In the optical-mechanical analogy the range r plays the role of time while the ray travel time t is an analog to Hamilton's principal function.[4,7,12,20]

B. Action-angle variables

Consider an *unperturbed* $(\delta c = 0)$ waveguide with Hamiltonian

$$H_0 = \frac{p^2}{2} + U_0(z).\tag{8}$$

For simplicity we assume that $U_0(z)$ has a single minimum. In this (range-independent) waveguide each ray is periodic and the Hamiltonian H_0 remains constant along the trajectory. The action variable I is defined by an integral over a cycle of the ray path[7,20]

$$I = \frac{1}{2\pi}\oint p\,dz = \frac{1}{\pi}\int_{z_{\min}}^{z_{\max}} dz\sqrt{2[H_0 - U_0(z)]},\tag{9}$$

where $z_{\min}$ and $z_{\max}$ are the lower and upper turning depths, respectively, satisfying the condition $U_0(z) = H_0$. Equation (9) determines Hamiltonian H_0 as a function of action I. Note that the action I grows with $|\chi_a|$, where χ_a is a ray grazing angle at the sound channel axis. Therefore steep rays have greater actions than the flat ones. The derivative

$$\frac{dH_0(I)}{dI} = \omega(I)\tag{10}$$

is a spatial frequency of trajectory oscillations.[3,7]

The momentum-position, (p,z), and action-angle, (I,θ), variables are connected by a well-known canonical transformation[7,20,22]

$$p = \frac{\partial G}{\partial z}, \quad \theta = \frac{\partial G}{\partial I},\tag{11}$$

where

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Virovlyansky *et al.*: Statistical description of chaotic rays   2543

$$G(I,z) = \begin{cases} \int_{z_{\min}}^{z_{\max}} dz\sqrt{2[H_0(I) - U_0(z)]}, & p > 0 \\ 2\pi - \int_{z_{\min}}^{z_{\max}} dz\sqrt{2[H_0(I) - U_0(z)]}, & p < 0 \end{cases}$$

$$(12)$$

is a generating function. Equations (9), (11), and (12) define the direct

$$I = I(p,z), \quad \theta = \theta(p,z) \tag{13}$$

and inverse

$$p = p(I,\theta), \quad z = z(I,\theta) \tag{14}$$

canonical transformations. According to our definition of generating function G the ray cycle begins at a minimum of the ray path and ends up at the next one. The angle variable θ at the ray cycle varies from 0 to 2π.

The Liouville theorem reads[20]

$$\frac{\partial(I(p,z), \theta(p,z))}{\partial(p,z)} = \frac{\partial(p(I,\theta), z(I,\theta))}{\partial(I,\theta)} = 1. \tag{15}$$

Although our canonical transformation is determined on the basis of relations connecting parameters of ray paths in an unperturbed waveguide, formally it can be used in a perturbed waveguide ($\delta c \neq 0$) as well. In the presence of perturbation the ray equations take the form

$$\frac{dI}{dr} = -\frac{\partial V}{\partial \theta}, \tag{16}$$

and

$$\frac{d\theta}{dr} = \omega + \frac{\partial V}{\partial I}. \tag{17}$$

As has been indicated in Sec. I, the angle variable θ can be interpreted as a phase of the ray path. At the beginning of a cycle θ takes value $\theta = 0$, passes through π at the maximum, and reaches 2π at the end of cycle. Then it jumps again to $\theta = 0$ and all repeats. To make θ continuous, its value should be increased by 2π at the beginning of each new cycle. This is a standard procedure.[20] Then both functions on the right-hand side of Eq. (14) are periodic in θ with period 2π.

C. Environmental model

In numerical simulations presented in the following we use an environmental model with an unperturbed profile $c_0(z)$ shown in the left panel of Fig. 1. This $c_0(z)$ (borrowed from Ref. 23) represents the Munk profile[5,24]

$$c_0(z) = c_r[1 + \varepsilon(e^\eta - \eta - 1)], \quad \eta = 2(z - z_a)/Q \tag{18}$$

with parameters $c_r = 1.5$ km/s, $\varepsilon = 0.00238$, $Q = 0.485$ km, and $z_a = -0.7$ km.

It is assumed that the weak perturbation $\delta c(r,z)$ is caused by random internal waves with statistics determined by the empirical Garrett-Munk spectrum.[5] To generate real-
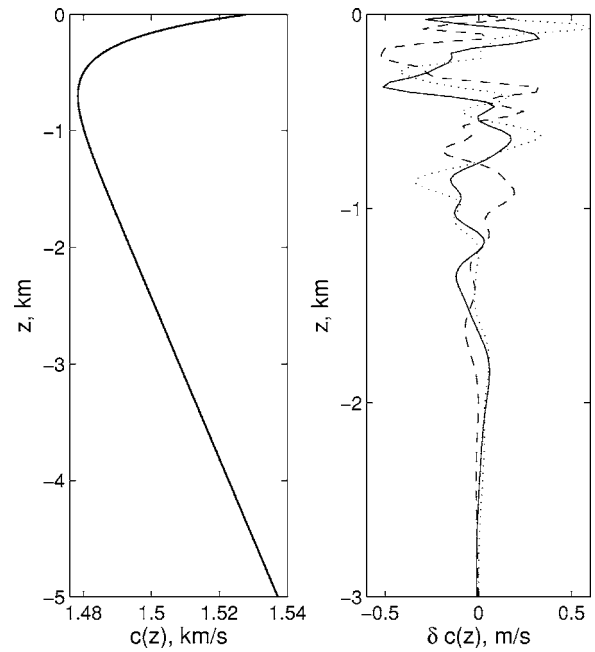
FIG. 1. Unperturbed sound speed profile $c_0(z)$ (left panel) and perturbation δc in vertical sections of the waveguide at three different ranges (right panel).

izations of a random field $\delta c(r,z)$ we apply a numerical technique developed by Colosi and Brown.[25] In their model the perturbation has the form

$$\delta c(r,z) = c_r \frac{\mu}{g} N^2 \zeta(r,z), \tag{19}$$

where $g = 9.8$ m/s^2 is the gravitational acceleration, $\mu = 24.5$ is a dimensionless constant, $N(z) = N_0 \exp(z/L)$ is a buoyancy frequency profile, $N_0 = 2\pi/(12 \text{ min}) = 0.0087$ 1/s is a buoyancy frequency near the surface, and $L = 1$ km. The random function $\zeta(r,z)$ presents internal-wave-induced vertical displacements of a fluid parcel. Its realizations have been computed using Eq. (19) from Ref. 25. We consider an internal wave field formed by 30 normal modes and assume its horizontal isotropy. Components of wave number vectors in the horizontal plane belong to the interval from $2\pi/100$ km^{-1} to $2\pi/4$ km^{-1}. A rms amplitude of the perturbation scales in depth like $\exp(3z/2L)$ and its surface-extrapolated value in our model is about 0.5 m/s. Depth dependencies of δc at three different ranges are shown in the right panel of Fig. 1.

III. STATISTICAL DESCRIPTION OF CHAOTIC RAYS IN A DEEP SEA

A. Probability density functions of ray parameters

In this work we consider a statistical description of chaotic rays based on a property of the chaotic ray dynamics called *mixing*.[10,11] Take a bundle of rays starting from a small area of the phase space $\mathcal{R}$ centered at a point (p_0, z_0). The square of $\mathcal{R}$ denote by $S_\mathcal{R}$. At ranges $r \gg \beta^{-1}$ points depicting the ray paths in the phase plane are scattered over much larger area $\mathcal{R}'$. Consider a small portion of this area $\Delta\mathcal{R}'$ centered at an observation point (p,z). A fraction of trajec-

tories arriving at $\Delta\mathcal{R}'$ may be treated as an expectancy of hitting this area. In order to quantify this statement, introduce a function

$$P_{pz|p_0 z_0}(p,z,r|p_0,z_0) = \frac{1}{S_\mathcal{R}} \int \int_\mathcal{R} dp_0' dz_0' \delta(z$$
$$- z(r,p_0',z_0',0)) \delta(p - p(r,p_0',z_0',0)) \quad, \tag{20}$$

which may be interpreted as a conditional probability density function (PDF) of ray coordinates in the phase plane. Then an integral over the area $\Delta\mathcal{R}'$,

$$P_{\Delta\mathcal{R}'} = \int_{\Delta\mathcal{R}'} dp\, dz\, P_{pz|p_0 z_0}(p,z,r-r_0|p_0,z_0), \tag{21}$$

determines the probability that a ray path starting from (p_0,z_0) arrives at $\Delta\mathcal{R}'$. The mixing phenomenon manifests itself in the fact that at long ranges $P_{\Delta\mathcal{R}'}$ weakly depends on the size and shape of $\mathcal{R}$. When r tends to infinity and sizes of $\mathcal{R}$ tend to zero the dependence of $P_{\Delta\mathcal{R}'}$ on the shape of $\mathcal{R}$ vanishes. But at finite ranges r we cannot take too small $\mathcal{R}$'s and consider probabilities of hitting too small $\Delta\mathcal{R}'$'s.

In a similar way we define a conditional PDF $P_{I\theta|I_0\theta_0}(I,\theta,r|I_0,\theta_0)$ for action-angle variables. The connection between PDFs $P_{pz|p_0 z_0}$ and $P_{I\theta|I_0\theta_0}$ is readily established using standard formulas of the probability theory. Equations (13) and (14) formally determine a nonlinear change of variables. The Liouville theorem (15) simplifies the application of standard relations and we arrive at

$$P_{pz|p_0 z_0}(p,z,r|p_0,z_0)$$
$$= P_{I\theta|I_0\theta_0}(I(p,z),\theta(p,z),r|I(p_0,z_0),\theta(p_0,z_0)). \tag{22}$$

The PDF introduced by Eq. (20) can be used for any environmental model in which rays exhibit chaotic behavior. But the application of this approach in a waveguide representing a realization of a random medium may have the following specific feature. At long ranges initially close ray paths diverge so significantly that they are spaced apart from each other by intervals exceeding correlation scales of the medium. Then the rays travel through practically independent inhomogeneities and behave as if they propagate in different realizations of the medium. Therefore it is natural to expect that the averaging over initial conditions may give results close to those obtained by the ensemble averaging. It means that the PDF defined by Eq. (20) may comparatively weakly depend on a particular realization of the waveguide. Numerical simulations show that this is the case for our environmental model (see the following).

B. Langevin stochastic equations for action and angle

As has been indicated in Sec. I, the angle variable rapidly randomizes and when studying wave fields at multi-megameter ranges, θ modulo 2π can be considered as a random variable independent of I and uniformly distribution on $(0,2\pi)$. Then the joint PDF of I and θ can be presented in the form

$$P_{I\theta|I_0\theta_0}(I,\theta,r|I_0,\theta_0) = \frac{1}{2\pi} P_{I|I_0}(I,r|I_0). \tag{23}$$

In Refs. 17 and 18 it has been shown that the PDF $P_{I|I_0}(I,r|I_0)$ describing the diffusion of action may be approximately evaluated analytically. Our approach is based on the fact that due to the weakness of perturbation δc the variation of action at a longitudinal correlation scale, l_r, is small. Therefore the right-hand side of Eq. (16) can be formally considered as a delta-correlated random function. Then $I(r)$ represents a Markov process whose PDF $P_{I|I_0}$ obeys the Fokker-Planck equation[10,11,21]

$$\frac{\partial P_{I|I_0}}{\partial r} = \frac{\partial}{\partial I} B(I) \frac{\partial P_{I|I_0}}{\partial I}. \tag{24}$$

In order to estimate the diffusivity $B(I)$ we have exploited a numeric procedure discussed in Refs. 17 and 18. It turns out that in the model under consideration the diffusivity weakly depends on I and it can be approximated by constant

$$B = 1.4 \times 10^{-7} \text{ km}. \tag{25}$$

In this approximation the range dependence of action is presented as

$$I(r) = I_0 + x(r), \tag{26}$$

where $I_0 \equiv I(0)$ and $x(r)$ is the so-called Wiener process.[21] The latter is defined by a stochastic Langevin equation

$$\frac{dx}{dr} = \xi(r), \quad x(0) = 0, \tag{27}$$

where $\xi(r)$ is a zero mean white noise with a correlation function

$$\langle \xi(r)\xi(r')\rangle = B\delta(r-r'). \tag{28}$$

The PDF of the Wiener process is well known:[21]

$$P_x(x) = \frac{1}{\sqrt{2Br}} \exp\left(-\frac{x^2}{2Br}\right). \tag{29}$$

There is a subtlety to this result. According to Eqs. (26) and (29) action I may take on both positive and negative values. But the action is non-negative by definition. In order to overcome this contradiction we introduce a reflecting boundary for the Wiener process at $x=-I_0$. This is an additional assumption whose applicability is justified by numerical simulation. From the viewpoint of Fokker-Planck equation the presence of the boundary is accounted for by condition

$$\left. \frac{\partial P_{I|I_0}}{\partial I} \right|_{I=0} = 0. \tag{30}$$

A solution to the Fokker-Planck equation (24) with an initial condition $P_{I|I_0}(I,0|I_0) = \delta(I-I_0)$ and boundary condition (30) is[21]

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Virovlyansky *et al.*: Statistical description of chaotic rays   2545

$$P_{I|I_0}(I, r|I_0) = \frac{1}{\sqrt{2Br}}\left[\exp\left(-\frac{(I-I_0)^2}{2Br}\right)\right.$$

$$\left. + \exp\left(-\frac{(I+I_0)^2}{2Br}\right)\right]. \tag{31}$$

This gives explicit expressions for the mean deviation of action from its starting value,

$$\langle I - I_0 \rangle = +\sqrt{\frac{2Br}{\pi}}\exp\left(-\frac{I_0^2}{2Br}\right) - I_0\,\mathrm{erfc}\left(\frac{I_0}{\sqrt{2Br}}\right), \tag{32}$$

and the dispersion of this deviation,

$$\langle (I - I_0)^2 \rangle = Br - 2\sqrt{\frac{2Br}{\pi}}I_0\exp\left(-\frac{I_0}{2Br}\right)$$

$$+ 2I_0^2\,\mathrm{erfc}\left(\frac{I_0}{\sqrt{2Br}}\right), \tag{33}$$

where

$$\mathrm{erfc}(u) = \frac{2}{\sqrt{\pi}}\int_u^\infty e^{-v^2}dv \tag{34}$$

is a complementary error function.[26]

For steep rays with I_0 satisfying condition

$$I_0 > (Br)^{1/2} \tag{35}$$

the second exponential on the right-hand side of Eq. (31) may be neglected. Most part of such rays have not reflected off the boundary $I=0$. Then $\langle I \rangle = I_0$ and

$$\sigma_I \equiv \langle (I-I_0)^2 \rangle^{1/2} = (Br)^{1/2}. \tag{36}$$

In numerical examples presented in the following we shall study statistics of chaotic rays at $r = 3000$ km. At this range $(Br)^{1/2} = 0.02$ km, which is approximately equal to I_0 of a ray whose grazing angle at the sound channel axis χ_a is close to 5°. Rays with $\chi_a > 5°$ will be called the steep ones. Rough analytical estimates for description of their distributions can be obtained using Eq. (31) without the second exponential on the right.

We now turn our attention to the derivation of a stochastic Langevin equation for the angle variable θ. We shall consider an "unwrapped" θ monotonically increasing with range (see the comment at the end of Sec. II B). Let us present it in the form

$$\theta = \theta_0 + \omega(I_0)r + y(r), \tag{37}$$

where $y(r)$ is a difference between angle variables of perturbed and unperturbed paths with equal starting parameters $I(0) = I_0$ and $\theta(0) = \theta_0$. Using an approximation $\omega(I) = \omega(I_0) + \omega'(I_0)x$ and neglecting the last term on the right of Eq. (17) yields[17]

$$\frac{dy}{dr} = \omega'(I_0)x. \tag{38}$$

According to this equation

$$y(r) = \omega'(I_0)\int_0^r x(r')dr', \tag{39}$$

that is the random function $y(r)$ is proportional to an integral of the Wiener process. For steep rays satisfying condition (35) we have $\langle y \rangle = 0$, and

$$\sigma_\theta \equiv \langle y^2 \rangle^{1/2} = |\omega'(I_0)|(B/3)^{1/2}r^{3/2}. \tag{40}$$

A statistical approach for the description of chaotic rays based on Langevin equations (27) and (38) we shall call the Wiener process approximation.

Analytical evaluation of statistical moments of y in the presence of reflecting boundary at $x = -I_0$ is tedious. But this problem can be easily solved numerically. To this end, we approximately replace the continuous random function $x(r)$ by a sequence of random variables x_m associated with discrete range points $r_m = m\varepsilon$, where $m = 0, \ldots, N$, and $\varepsilon = r/N$. Values of x_m are evaluated using a recurrence formula

$$x_{m+1} = x_m + \xi_m, \quad x_0 = 0, \tag{41}$$

where ξ_n are zero mean Gaussian random variables (produced by a standard generator of random numbers) with correlation function

$$\langle \xi_{m_1}\xi_{m_2} \rangle = \frac{B}{\varepsilon}\delta_{m_1 m_2}. \tag{42}$$

The presence of reflecting boundary is accounted for in the following way. At each step the value of x_{m+1} is compared with $-I_0$. If x_{m+1} turns out to be less than $-I_0$, it is replaced by $-2I_0 - x_{m+1}$. Simple numerical tests confirm that at $N \gg 1$ the PDF of the random variable $I = I_0 + x_N$ is well described by function (31). Values of random function $y(r)$ at points r_m are computed by

$$y_m = \omega'(I_0)\sum_{q=1}^m x_q\varepsilon. \tag{43}$$

In what follows we shall use Eqs. (41)–(43) to model realizations of $x(r)$ and $y(r)$ and evaluate their PDFs and statistical moments at a range of 3000 km. These calculations will be done with $\varepsilon = 10$ km.

IV. PROBABILITY DENSITY FUNCTIONS OF RAY PARAMETERS

To test the validity of the Wiener process approximation in the following we compare its predictions with results obtained by direct numerical ray tracing.

A. Rays starting from a small area in the phase space

Our analytic description is based on Eq. (31), which predicts a distribution of actions for rays starting from a small area $\mathcal{R}$ in the phase plane. Results of numerical ray tracing carried out to verify this formula are presented in Figs. 2(a)–2(f). We have selected three areas $\mathcal{R}$ shown in Figs. 2(a), 2(c), and 2(e) by closed rectangles. The rectangle in Fig. 2(a) defines rays with initial depths close to the sound channel axis z_a and launch angles centered at 3°. 10 000 rays starting from points (p_0, z_0) uniformly filling this rectangle
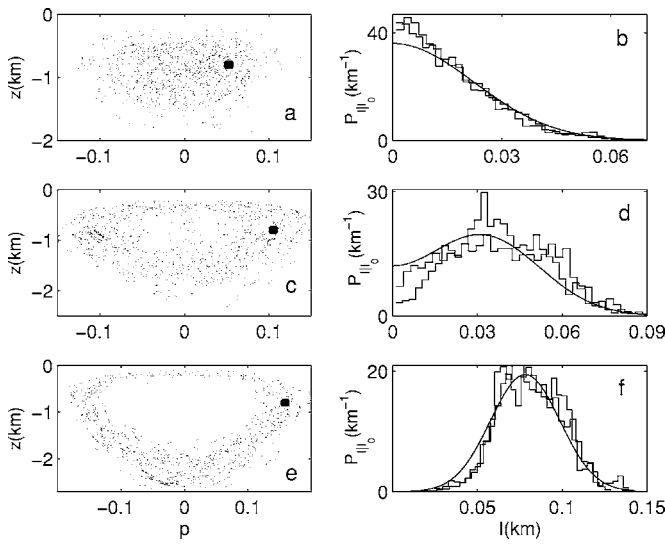
FIG. 2. Distributions of rays escaping a small area $\mathcal{R}$ in the phase plane. Three selected areas $\mathcal{R}$ are shown in plots (a), (c), and (e) by small closed rectangles. Randomly scattered points in each of these plots depict coordinates of rays (starting from a corresponding rectangle) at a range of 3000 km. The right column of plots present distributions of action I at 3000 km. Stairstep graphs show normalized histograms computed by direct ray tracing for two different realizations of $\delta c(r,z)$. Smooth curves show PDF predicted by Eq. (31). Starting depths of rays escaping all three rectangles are close to the sound channel axis. Launch angles are close to $3°$ [(a) and (b)], $6°$ [(c) and (d)], and $9°$ [(e) and (f)].

have been traced up to $r=3000$ km. Randomly scattered points in Fig. 2(a) depict the distribution of these rays in the phase plane (p,z) at 3000 km. The distribution of actions of these rays is presented by one of two normalized histograms (stairstep graphs) in Fig. 2(b). Another histogram in Fig. 2(b) presents an analogous distribution computed for a different realization of $\delta c(r,z)$. The normalized histograms are compared to a PDF predicted by Eq. (31) (smooth curve). Similar results for rays with launch angles centered at $6°$ and $9°$ are presented in Figs. 2(c)–2(f).

In accord with our conjecture made at the end of Sec. III A the normalized histograms obtained for different $\delta c(r,z)$ look alike and they are in a reasonable agreement with estimates following from Eq. (31). Similar results have been observed for rays traced using a few other realizations of $\delta c(r,z)$ (not shown).

B. Rays escaping a point source

Take a point source set at a depth z_0 and assume that the quantity $\eta(p_0)dp_0$ determines a fraction of rays with starting momenta from an interval (p_0, p_0+dp_0). The function $\eta(p_0)$ is determined by the radiation pattern of a particular source. Formally considering $\eta(p_0)$ as a PDF of the initial momentum p_0 and exploiting Eqs. (22), (23), and (31) we can find PDFs of I, p, and z at ranges $r \gg \beta^{-1}$. Making use of Eq. (31) we get the PDF of action in the form

$$P_I(I,r) = \int dp_0\, \eta(p_0) P_{I|I_0}(I,r|I(p_0,z_s)). \tag{44}$$

Exploiting standard formulas of the probability theory and the Liouville theorem (15), from Eq. (23) we get a joint PDF of p and z:
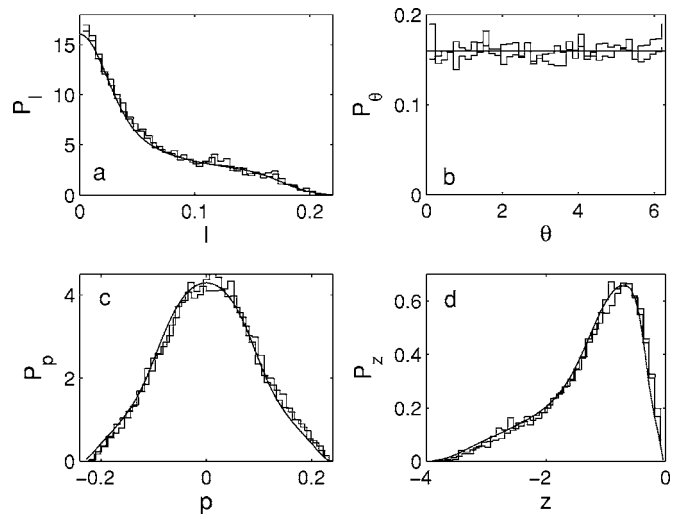
FIG. 3. Distributions of ray parameters at $r=3000$ km. Smooth curves show PDFs of I (a), θ (b), p (c), and z (d) evaluated analytically in the Wiener process approximation for rays escaping a point source set at a depth $z =-0.7$ km. Stairstep graphs present normalized histograms obtained by numerical ray tracing in waveguides with two different realizations of perturbation $\delta c(r,z)$.

$$P_{pz}(p,z,r) = \frac{1}{2\pi} P_I(I(p,z),r). \tag{45}$$

Equations (44) and (45) yield a PDF of momentum

$$P_p(p,r) = \frac{1}{2\pi} \int dp_0 dz\, \eta(p_0) P_{I|I_0}(I(p,z),r|I(p_0,z_s)), \tag{46}$$

and depth

$$P_z(z,r) = \frac{1}{2\pi} \int dp_0 dp\, \eta(p_0) P_{I|I_0}(I(p,z),r|I(p_0,z_s)). \tag{47}$$

To check the above-mentioned relations we have traced 50 000 rays escaping a source set at depth $z_0=z_a$ with initial momenta uniformly distributed in an interval $(-p_{max}, p_{max})$, with $p_{max}=0.22$ corresponding to a maximum launch angle $\chi_{max}=12.5°$. In this case $\eta(p_0)=(2p_{max})^{-1}$. Smooth lines in Fig. 3 show PDFs of I (a), θ (b), p (c), and z (d) at a range of 3000 km obtained in the Wiener process approximation, that is evaluated by formulas (44), (46), and (47). For the angle variable θ modulo 2π the uniform distribution is expected. The smooth curves are compared to stairstep graphs representing distributions (normalized histograms) of the same quantities obtained by direct numerical ray tracing for two different realizations of $\delta c(r,z)$. It is clear that predictions made in the Wiener process approximation are in good agreement with results of ray tracing. In these figures we see again that distributions of ray parameters weakly depend on particular realizations of perturbation.

V. STATISTICS OF RAY TRAVEL TIMES

In Refs. 17 and 27 approximate analytical relations connecting variations (caused by internal waves) of geometrical characteristics of the ray path, I, θ, p, and z, to variations of ray travel times were derived. Combining these relations and

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Virovlyansky *et al.*: Statistical description of chaotic rays   2547

Eq. (31) yields a statistical description of ray travel times.[17,19] This issue is discussed in the following.

A. Variations of ray travel times caused by internal waves

First, consider a perturbed and unperturbed ray escaping a point source at equal launch angles. Initial action and angle variables of these rays are denoted by I_0 and θ_0, respectively. In Refs. 17, 19, and 27 it has been shown that even at ranges $O(1000 \text{ km})$ the difference in travel times of these rays, Δt, can be approximately estimated analytically as

$$\Delta t = \tau_G + \tau_N + \tau_I + \tau_V, \tag{48}$$

where

$$\tau_G = [G(z,I_0) - G(\bar{z},I_0)]/c_r, \tag{49}$$

z and $\bar{z}$ are coordinates of perturbed and unperturbed rays, respectively, at the observation range;

$$\tau_N = 2\pi \Delta N I_0/c_r, \tag{50}$$

ΔN is the difference between numbers of minima (numbers of cycles) of perturbed and unperturbed ray paths;

$$\tau_I = \frac{\omega'(I_0)}{2c_r} \int_0^r [I(r') - I_0]^2 dr', \tag{51}$$

and

$$\tau_V = -\frac{1}{c_r} \int_0^r V(I_0, \theta(r'), r') dr. \tag{52}$$

Functions $I(r)$ and $\theta(r)$ in Eqs. (51) and (52) present the range dependencies of action and angle variables, respectively, along the perturbed ray path.

At comparatively short ranges $O(100 \text{ km})$ both rays follow practically the same path. Then $\Delta N = 0$, $z \simeq \bar{z}$, $I(r') \simeq I_0$ and the terms τ_G, τ_N, and τ_I on the right-hand side of Eq. (48) are negligible. Equation (48) reduces to a well-known relation,[5]

$$\Delta t \simeq \tau_V \simeq -\frac{1}{c_r^2} \int_0^r \delta c(r,z_0(r)) dr, \tag{53}$$

where the integration goes along the unperturbed path $z_0(r)$. In this case the travel time variations are caused by sound speed fluctuations crossed by the unperturbed ray. A detailed analysis of statistical characteristics of τ_V have been done in Ref. 5. The rms variations of τ_V grows with range as $r^{1/2}$.

At ranges $O(1000 \text{ km})$ perturbed and unperturbed paths with equal starting parameters are no longer close and usually have different numbers of cycles $(\Delta N \neq 0)$.[3,17,19,27] For steep rays with large starting actions I_0 the main contribution to Δt at long ranges comes from the term τ_N.[17,19,27] From Eqs. (9) and (12) it follows that $2\pi I_0 > G(I_0,z)$ and, hence, $\tau_G < \tau_N$. Therefore the term τ_N may be used as rough estimates of Δt for steep rays. For flat rays (small I_0) both τ_N and τ_G become negligible as compared to τ_I. Since the latter typically exceeds τ_V, the relation
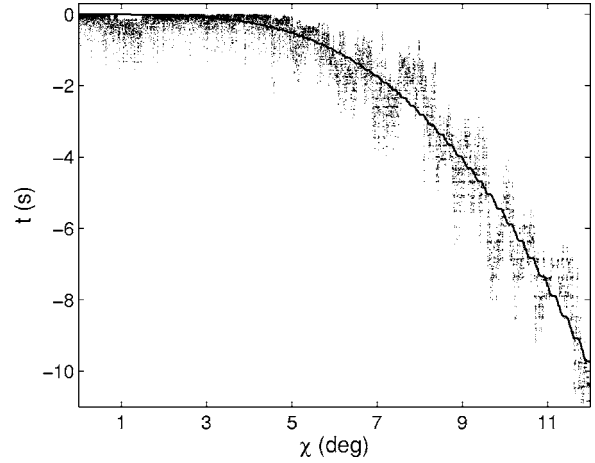
FIG. 4. Ray travel time vs launch angle in the unperturbed (solid line) and perturbed (points) waveguides at a range of 3000 km. Rays escape a point source set at a depth $z = -0.7$ km.

$$\Delta t = \tau_N + \tau_I \tag{54}$$

gives a rough estimate of Δt at ranges $O(1000 \text{ km})$ for rays with arbitrary launch angles.

To apply the Wiener process approximation for description of Δt we approximate ΔN by $y/(2\pi)$. This yields

$$\tau_N = y(r) I_0/c_r. \tag{55}$$

According to Eqs. (26) and (51),

$$\tau_I = \frac{\omega'(I_0)}{2c_r} \int_0^r x^2(r') dr'. \tag{56}$$

Equations (54)–(56) allow one to model the distribution of Δt using the procedure described at the end of Sec. III B. Then the dependence of Δt on the launch angle enters through the initial action I_0.

Figure 4 shows the ray travel time at 3000 km as a function of launch angle in the unperturbed (smooth curve) and perturbed (points) waveguide. Perturbed travel times have been computed using the same fan of rays as in Sec. IV B. Note that a similar dependence of the ray travel time on the launch angle is observed in a waveguide with periodic range dependence. Some characteristic features of this dependence were investigated in Refs. 28 and 29.

To find unperturbed travel times, a fan of rays with the same starting parameters have been traced in the unperturbed waveguide. The difference in travel times of perturbed and unperturbed rays versus launch angle is shown in the upper panel of Fig. 5. In Figs. 4 and 5 we consider rays with positive launch angles. Results for negative launch angles look similar (not shown).

Making use of Eqs. (41) and (43), we have evaluated 50 000 realizations of $x(r)$ and $y(r)$ for 3000 km propagation. Each pair $[x(r), y(r)]$ models a chaotic ray with starting action I_0. A set of ξ_n's for each ray has been generated independently. Then, by formulas (55) and (56) we have computed τ_N and τ_I for all rays. The lower panel in Fig. 5 presents the sum $\tau_N + \tau_I$ as a function of launch angle. It is seen that our simple model based on Eqs. (41)–(43) and (54)–(56) gives a reasonable prediction for the spread of Δt.
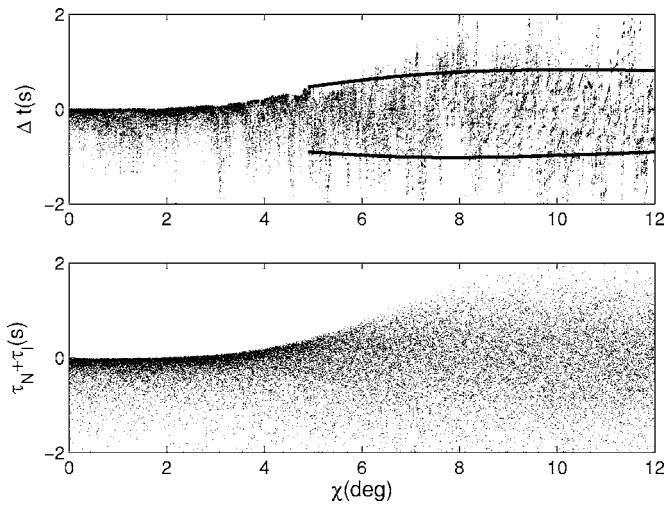
FIG. 5. Difference between travel times of perturbed and unperturbed rays with equal launch angles χ. Upper panel: Result of direct numerical ray tracing; solid lines show deviations $\pm\sigma_{\Delta t}$ predicted by Eq. (57). Lower panel: Simulation of sum $\tau_N + \tau_I$ in the Wiener process approximation on the basis of Eqs. (41), (43), (55), and (56).

Note that at the 3000 km range the term $\tau_V = O(5 \text{ ms})$ [5,17] and it is negligible compared to the spread of Δt seen in Fig. 5.

An analytical estimate for the rms value of Δt can be easily derived for steep rays. In this case $\Delta t \simeq \tau_N$ and $<\Delta t> = 0$. For steep rays we can ignore the reflecting boundary at $x = -I_0$ and exploit Eq. (40). Combining Eqs. (40) and (55) yields a standard deviation of Δt,

$$\sigma_{\Delta t} = |\omega'(I_0)| \frac{r^{3/2}}{c_r} \left(\frac{B}{3}\right)^{1/2}. \qquad (57)$$

Two thick solid lines in Fig. 5 depict $\pm\sigma_{\Delta t}$ versus launch angle. Equation (57) estimates the spread of travel times for a bundle of rays escaping a point source at (steep) launch angles from a narrow angular interval centered at angle χ_0 corresponding to the starting action I_0. According to Eq. (57) the rms spread of travel times grows with range like $r^{3/2}$. The same range dependence was found by other authors.[30]

B. Segments of the timefront

Figure 6 shows another representation of ray travel times at 3000 km. This is the so-called timefront: ray arrivals in the time-depth plane. The upper and lower panels in Fig. 6 present timefronts in the unperturbed and perturbed waveguide, respectively. The timefront in the unperturbed waveguide has the well-known accordion-like shape consisting of smooth segments (branches).[3,4,14,24] Each segment is formed by rays with the same identifier $\pm M$, where $\pm$ is the sign of launch angle and M is the number of ray turning points.

In the perturbed timefront, points depicting arrivals of rays with the given identifier are scattered in the vicinity of the corresponding unperturbed segment. These points form what we call the fuzzy segment. Fuzzy segments in the early part of the received pattern are formed by steep rays. They do not overlap and are very close to the unperturbed ones with corresponding identifiers.

In the present section we study the spreads of points forming the fuzzy segment. Unperturbed and perturbed segments in Fig. 6 formed by rays with identifier +140 are marked by thick line (upper panel) and thick points (lower panel). An expanded view of these segments is shown in Fig. 7: the thick solid line again presents the unperturbed segment and randomly scattered points depict the fuzzy one. Our task is to derive a quantitative estimate of the width of a fuzzy segment. Introduce a quantity τ_i (see Fig. 7) that represents a time delay between the arrival of ith ray contributing to the given fuzzy segment and the unperturbed segment corresponding to the same identifier. In fact, τ_i is a difference in travel times of perturbed and unperturbed rays escaping a point source and arriving at the same observation point $(z = \bar{z})$ with equal identifiers. Let us apply formulas (48)–(52) to estimate τ_i. Setting $\tau_G = \tau_N = 0$ and, neglecting τ_V, we get [17,19,27]

FIG. 6. Timefronts in the unperturbed (upper panel) and perturbed (lower panel) waveguide: Depth vs ray travel time at 3000 km for a point source set at a depth $z = -0.7$ km. Identifiers of rays forming some particular segments are indicated next to the corresponding segments. In the upper panel, arrivals with identifier +140 are depicted by a thick solid line. In the lower panel, arrivals with this identifier are marked by thick points.

Virovlyansky *et al.*: Statistical description of chaotic rays   2549

FIG. 7. Arrivals with identifiers +140 at the range 3000 km are shown in the time-depth plane. Points and a solid line depict arrivals with and without internal waves present, respectively.

FIG. 8. Mean bias of the fuzzy segment at the sound channel axis as a function of number of ray turning points, M. The plot is constructed for rays starting upward. Thin solid line: Direct ray tracing. Thick solid line: Analytical estimate (64). Dashed line: Simulation in the Wiener process approximation based on Eqs. (41), (43), and (60).

$$\tau_i = \frac{\omega'(\bar{I}_M)}{2c_r} \int_0^r [I(r') - \bar{I}_M]^2 dr', \tag{58}$$

where $I(r)$ is an action of the ith perturbed ray as a function of range, and $\bar{I}_M$ is the mean action of rays forming the unperturbed segment. According to Eq. (58) the sign of τ_i is determined by the sign of $\omega'(\bar{I}_M)$. For typical deep water waveguides the cycle length grows with launch angle $|\chi|$ and therefore $\omega'(I)$ is negative. Then $\tau_i < 0$ and the perturbed segment is biased toward early times.[17,19,27] Let us emphasize that this statement is valid only at long enough ranges where $|\tau_I|$ for most rays exceeds $|\tau_V|$. Note that at short ranges where τ_V dominates there exists a different mechanism of bias. It is discussed in Ref. 31.

Since all (or almost all) τ_i have the same sign, both width and bias of a fuzzy segment are characterized by the quantity

$$\delta\tau_{mean} = N^{-1} \sum_{i=1}^{N} \tau_i. \tag{59}$$

Summation is taken over rays belonging to the selected fuzzy segment and arriving at depths in an interval $z_{low} < z < z_{up}$, where z_{low} and z_{up} are depths of lower and upper ends of the unperturbed segment, respectively. A thin solid line in Fig. 8 shows $\delta\tau_{mean}$ for segments with identifiers $+M$ for $M = 110, 111, \ldots, 185$ obtained by direct ray tracing. Values of $\delta\tau_{mean}$ for segments with identifiers $+M$ and $-M$ are close for most M (not shown).

In order to estimate τ_i in the Wiener process approximation we present the action of the ith ray as $I(r) = I_{0,i} + x(r)$, where $I_{0,i}$ is a starting action. From Eq. (39) we get

$$\tau_i = \frac{\omega'(\bar{I}_M)}{2c_r} \left[(I_{0,i} - \bar{I}_M)^2 r + (I_{0,i} - \bar{I}_M) \int_0^r x(r')dr' \right.$$
$$\left. + \int_0^r x^2(r')dr' \right]. \tag{60}$$

To avoid confusion, note that in the preceding section we compared travel times of perturbed and unperturbed rays with equal starting parameters. This explains the difference between the right-hand sides of Eqs. (56) and (60), obtained on the basis of the same formula (51).

The fact that the rays under consideration have the same identifier can be (approximately) accounted for by introducing a constraint [17,19]

$$|\theta_i - \bar{\theta}_M| < \pi/2, \tag{61}$$

where θ_i and $\bar{\theta}_M$ are angle variables of the ith perturbed ray and the unperturbed ray with starting action $\bar{I}_M$, respectively. Making use of Eq. (37) we rewrite Eq. (61) as

$$|\omega(I_{0,i})r + y(r) - \omega(\bar{I}_M)r| < \pi/2. \tag{62}$$

Here we have neglected the difference in initial values of angle variables of perturbed and unperturbed rays. This additional approximation, that slightly relaxes the constraint (61), is justified by the fact that the values of $|y(r)|$ at 3000 km for most rays significantly exceed π. Employing a one more approximation $\omega(I_{0,i})r - \omega(\bar{I}_M) = \omega'(\bar{I}_M)(I_{0,i} - \bar{I}_M)$, we rewrite Eq. (62) in the form[17,19]

$$\left| I_{0,i} - \bar{I}_M + \int_0^r x(r')dr' \right| < \frac{\pi}{2\omega'(\bar{I}_M)}. \tag{63}$$

Using the numerical procedure described at the end of Sec. III B we have computed realizations of $x(r)$ for 50 000 fan rays. Selecting rays satisfying the constraint (63) and evaluating τ_i by formula (60) we have obtained desired estimates of $\delta\tau_{mean}$ shown in Fig. 8 by a dashed line.

Virovlyansky *et al.*: Statistical description of chaotic rays

For steep rays, statistical characteristics of τ_i, defined by Eq. (60) in the presence of constraint (63), can be evaluated analytically. In Refs. 17 and 19 it was shown that

$$\delta\tau_{\mathrm{mean}} = \frac{\omega'(\bar{I}_M)}{12c_r}Br^2. \tag{64}$$

In Fig. 8 this dependence for segments with $M < 160$ is displayed by a thick solid line. Segments with larger M are formed by flat rays. Figure 8 demonstrates that predictions (numerical and analytical) obtained in the Wiener process approximation agree fairly well with results of direct ray tracing.

VI. CONCLUSION

In this paper we have discussed an approximate statistical approach for description of chaotic ray dynamics in a deep water acoustic waveguide. The ray motion is studied using the Hamiltonian formalism. The Hamilton equations (16) and (17) for the action and angle variables are replaced by the stochastic Langevin equations (27) and (38) whose solutions are readily expressed through the random Wiener process. In Ref. 14 a similar approach was applied to the Hamilton equations for the momentum and position variables.

All our analytical estimates characterizing statistics of different ray parameters stem from Eq. (31). It is surprising that the PDF of action in a realistic environmental model may be properly described by so simple a formula. This results from the fact that in the model considered in the present paper (as well as in a slightly different model used in Refs. 17–19 and 27) the diffusivity $B(I)$ can be approximated by a constant. The latter summarizes in a single parameter all characteristics of perturbation defining the diffusion of action. In the general case $B(I) \neq \mathrm{const}$ and solving the Fokker-Planck equation (24) requires the use of numerical methods.

The canonical transformation connecting the pairs of variables (I, θ) and (p, z) is determined by the unperturbed sound speed profile. This transformation formally represents a nonlinear change of variables with a Jacobian equal to unity. Although statistical characteristics of p and z are expressed through those of I and θ using the standard formulas of the probability theory, the evaluation may be tedious, especially for flat rays when the approximation (35) cannot be exploited. Then it may be more convenient to apply an elementary numerical procedure described at the end of Sec. III B.

Statistics of ray travel times have been analyzed by combining the Wiener process approximation and analytical formulas (derived in Refs. 17 and 27) connecting travel time variations to variations of I and θ. Our statistical description makes it possible to get quantitative estimates characterizing the clusterization of travel times of chaotic rays. We have evaluated the bias and widening of the timefront segment formed by rays with the same identifier. These predictions are in a reasonable agreement with results of direct ray tracing. Less accurate predictions have been obtained for Δt representing the difference in travel times of perturbed and unperturbed rays with the same launch angle.

Numerical simulations support the conjecture formulated at the end of Sec. III A that statistical characteristics evaluated by averaging over ray starting parameters at long ranges weakly depend on a particular realization of perturbation $\delta c(r, z)$. Nevertheless, a question of how well a single realization of a random medium represents the whole statistical ensemble remains open and requires a further investigation.

In this paper we assume propagation in a vertical plane (r, z). It should be pointed out that if scattering in cross-range is allowed, some modification to the results presented can be expected. Useful analytical relations that may be helpful for taking into account the horizontal refraction of rays are obtained in Refs. 32 and 33.

Note another important issue that has not been broached here. We have not investigated the correlation between the travel time and ray amplitude. It is conceivable that, at a given observation point, the set of ray arrivals having the same identifier is dominated by a few arrivals while the rest have amplitudes below the detection threshold. In Ref. 3 it has been demonstrated numerically that the spread of travel times of the dominant rays may be significantly less than the spread of travel times corresponding to all arrivals.

[1] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, K. Metzger, M. Howe, A. Mercer, R. C. Spindel, J. A. Colosi, T. Birdsall, and A. B. Baggeroer, "A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern north Pacific ocean," J. Acoust. Soc. Am. **105**, 3185–3201 (1999).

[2] J. A. Colosi, E. K. Scheer, S. M. Flatte, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. Birdsall, and A. B. Baggeroer, "Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern north Pacific ocean," J. Acoust. Soc. Am. **105**, 3202–3218 (1999).

[3] F. J. Beron-Vera, M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in a long-range acoustic propagation experiment," J. Acoust. Soc. Am. **114**, 1226–1242 (2003).

[4] J. Simmen, S. M. Flatte, and G. Y. Wan, "Wavefront folding, chaos, and diffraction for sound propagation through ocean internal waves," J. Acoust. Soc. Am. **102**, 239–255 (1997).

[5] S. M. Flatte, R. Dashen, W. M. Munk, K. M. Watson, and F. Zakhariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, London, 1979).

[6] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarsky, *Introduction to Statistical Radiophysics. Part II* (Nauka, Moscow, 1978).

[7] S. S. Abdullaev and G. M. Zaslavsky, "Classical nonlinear dynamics and chaos of rays in wave propagation problems in inhomogeneous media," Usp. Fiz. Nauk **161**, 1–43 (1991).

[8] D. R. Palmer, M. G. Brown, F. D. Tappert, and H. F. Bezdek, "Classical chaos in nonseparable wave propagation problems," Geophys. Res. Lett. **15**, 569–572 (1988).

[9] D. R. Palmer, T. M. Georges, and R. M. Jones, "Classical chaos and the sensitivity of the acoustic field to small-scale ocean structure," Comput. Phys. Commun. **65**, 219–223 (1991).

[10]R. Z. Sagdeev, D. A. Usikov, and G. M. Zaslavsky, *Nonlinear Physics: From the Pendulum to Turbulence and Chaos* (Harwood Academic, Switzerland, 1988).

[11]A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, Applied Mathematical Sciences Vol. 38 (Springer, New York, 1992).

[12]L. E. Reichl, *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations* (Springer, New York, 1992).

[13]G. M. Zaslavsky, *Chaos in Dynamical Systems* (Harwood Academic, New York, 1985).

[14]M. G. Brown and J. Viechnicki, "Stochastic ray theory for long-range sound propagation in deep ocean environment," J. Acoust. Soc. Am. **104**, 2090–2104 (1998).

[15]K. B. Smith, M. G. Brown, and F. D. Tappert, "Ray chaos in underwater acoustics," J. Acoust. Soc. Am. **91**, 1939–1949 (1992).

[16]M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in long-range deep ocean sound propagation," J. Acoust. Soc. Am. **113**, 2533–2547 (2003).

[17]A. L. Virovlyansky, "Ray travel times in range-dependent acoustic waveguides," http://arxiv.org/abs/nlin/0012015 (2002). Viewed 4/3/2007.

[18]A. L. Virovlyansky, "Statistical description of ray chaos in an underwater acoustic waveguide," Acoust. Phys. **51**, 71–80 (2005).

[19]A. L. Virovlyansky, "Signal travel times along chaotic rays in long-range sound propagation in the ocean," Acoust. Phys. **51**, 271–281 (2005).

[20]L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed. (Pergamon, Oxford, 1976).

[21]C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, New York, 1985).

[22]S. S. Abdullaev, *Chaos and Dynamics of Rays in Waveguide Media*, edited by G. Zaslavsky (Gordon and Breach, New York, 1993).

[23]J. A. Colosi and S. M. Flatté, "Mode coupling by internal waves for multimegameter acoustic propagation in the ocean," J. Acoust. Soc. Am. **100**, 3607–3620 (1996).

[24]L. M. Brekhovskikh and Yu. P. Lysanov, *Fundamentals of Ocean Acoustics* (Springer, Berlin, 1991).

[25]J. A. Colosi and M. G. Brown, "Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation field," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).

[26]*Handbook of Mathematical Functions with Formulas, Graphs and Tables*, Applied Mathematical Series Vol. 55, edited by M. Abramovitz and I. A. Stegun (National Bureau of Standards, Washington, DC, 1964).

[27]A. L. Virovlyansky, "Ray travel times at long ranges in acoustic waveguides," J. Acoust. Soc. Am. **113**, 2523–2532 (2003).

[28]I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, "Sensitivity of ray travel times," Chaos **12**, 617–635 (2002).

[29]I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, "Ray chaos, travel time modulation, and sensitivity to the initial conditions," J. Acoust. Soc. Am. **117**, 1595–1606 (2005).

[30]F. J. Beron-Vera and M. G. Brown, "Travel time stability in weakly range-dependent sound channels," J. Acoust. Soc. Am. **115**, 1068–1077 (2004).

[31]J. L. Codona, D. B. Creamer, S. M. Flatté, R. G. Frehlich, and F. S. Henyey, "Average arrival time of wave pulses through continuous random media," Phys. Rev. Lett. **55**, 9–12 (1985).

[32]O. A. Godin, "Systematic distortions of signal propagation times in random inhomogeneous media," Dokl. Phys. **48**, 389–394 (2003).

[33]O. A. Godin, V. U. Zavorotny, A. G. Voronovich, and V. V. Goncharov, "Refraction of sound in a horizontally inhomogeneous, time-dependent ocean," IEEE J. Ocean. Eng. **31**, 384–401 (2006).

Frequency dependencies of phase velocity and attenuation coefficient in a water-saturated sandy sediment from 0.3 to 1.0 MHz

Kang Il Lee
Department of Physics, Kangwon National University, Chuncheon 200–701, Republic of Korea

Victor F. Humphrey
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

Byoung-Nam Kim and Suk Wang Yoon[a)]
Department of Physics and Institute of Basic Science, Sungkyunkwan University, Suwon 440-746, Republic of Korea

The frequency-dependent phase velocity and attenuation coefficient for the fast longitudinal wave in a water-saturated sandy sediment were measured over the frequency range from 0.3 to 1.0 MHz. The experimental data of phase velocity exhibited the significant negative dispersion, with the mean rate of decline of 120 ± 20 m/s/MHz. The Biot model predicted the approximately nondispersive phase velocity and the grain-shearing (GS) model exhibited the slightly positive dispersion. In contrast, the predictions of the multiple scattering models for the negative dispersion in the glass-grain composite were in general agreement with the experimental data for the water-saturated sandy sediment measured here. The experimental data of attenuation coefficient was found to increase nonlinearly with frequency from 0.3 to 1.0 MHz. However, both the Biot and the GS models yielded the attenuation coefficient increasing almost linearly with frequency. The total attenuation coefficient given by the algebraic sum of absorption and scattering components showed a reasonable agreement with the experimental data for overall frequencies. This study suggests that the scattering is the principal mechanism responsible for the variations of phase velocity and attenuation coefficient with frequency in water-saturated sandy sediments at high frequencies.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2713690]

I. INTRODUCTION

The Biot model (Biot, 1956a, b, 1962) for elastic wave propagation in porous media has been applied to sediment acoustics with varying degrees of success. It was originally applied to the analysis of ultrasound geophysical test data for porous rock samples. In 1970, the model was applied by Stoll and Bryan (1970) in a form more suitable for water-saturated sediments. Recently, Buckingham (1997, 2000, 2005) developed the grain-shearing (GS) model for wave propagation in saturated, unconsolidated granular materials, including marine sediments. Since the mineral grains are unbonded, it is assumed that the shear rigidity modulus of the medium is zero, implying the absence of a skeletal elastic frame. The GS model has been successfully applied to predict the relationship between mechanical properties, such as grain size and porosity, and acoustic properties of marine sediments. Chotiros and Isakson (2004) also proposed a new theoretical model based on the grain-to-grain contacts in unconsolidated granular materials.

Dispersion of phase velocity has been extensively investigated in marine sediments by a number of authors (Hamp-ton, 1967; Wingham, 1985; Turgut and Yamamoto, 1990; Stoll, 2002; Tang *et al.*, 2002; Williams *et al.*, 2002; Chotiros and Isakson, 2004). Most of the studies have been performed over various frequency ranges from a few hundreds of hertz up to a few hundreds of kilohertz, where the propagation of interest occurs over ranges that are sufficiently long to include bottom interactions. The fast longitudinal wave in marine sediments has been known to exhibit the positive dispersion of phase velocity at frequencies lower than about 200 kHz, which is predicted by theoretical models such as the Biot and the GS models. In contrast, an interesting high-frequency experimental study demonstrated that the fast wave velocity in water-saturated sandy sediments shows the negative dispersion from 0.2 to 1.2 MHz (Moussatov *et al.*, 1998). However, the velocity dispersion predicted by the Biot model has been reported to be very weak, but positive at frequencies higher than 0.2 MHz (Moussatov *et al.*, 1998), and the GS model also tends to predict the slightly positive dispersion in this frequency range (Buckingham, 1997). These results may imply that the underlying mechanism responsible for the velocity dispersion in the high frequency range is different from that in the low frequency range. Thus,

---

a)Electronic mail: swyoon@skku.ac.kr

both the Biot and the GS models may only have validity in a limited range of frequency for the dispersion of phase velocity in water-saturated sediments.

As extensively discussed by Buckingham (2005) and Ohkawa (2006), the frequency dependence of attenuation coefficient in marine sediments has been argued for many years. The Biot model predicts an attenuation coefficient increasing as the half power of frequency, which is due to the fluid viscosity only. Nolle *et al.* (1963) found the $f^{1/2}$ dependence of attenuation coefficient in water-saturated sands at frequencies of 0.2, 0.5, and 1.0 MHz. In contrast, most data exhibit a linear scaling of attenuation coefficient with frequency over a wide range of frequencies (Wingham, 1985; Hamilton, 1987), which is consistent with the prediction of the GS model based on grain-to-grain shearing mechanisms (Buckingham, 1997). In fact, the GS model yields an attenuation coefficient that is almost but not quite linear in frequency. On the other hand, Ohkawa (2006) demonstrated that the frequency dependence of attenuation data collected during the sediment acoustics experiment in 1999 (SAX99) follows $f^{1/2}$ at frequencies below 50 kHz, as the Biot model predicts, and the deviation of attenuation from the Biot model at frequencies higher than 200 kHz is due to effects of scattering. This suggests that the scattering is likely to be the principal mechanism responsible for the attenuation in the ultrahigh frequency range.

The present study aims to provide an insight into the frequency dependencies of phase velocity and attenuation coefficient for the fast longitudinal wave in a water-saturated sandy sediment over the frequency range from 0.3 to 1.0 MHz. The frequency-dependent phase velocity and attenuation coefficient in the water-saturated sandy sediment were measured using two different matched pairs of transducers with a diameter of 25.4 mm and center frequencies of 0.5 and 1.0 MHz. They were compared with the predictions obtained from the Biot and the GS models.

II. MATERIALS AND METHODS

A. Sandy sediment

The sandy sediment used in the present study was composed of clean medium sand with a porosity of 0.408 ± 0.013 and a mean grain diameter of 425 ± 84 μm spanned from 250 to 500 μm. The grain sizes of the sediment were much smaller than the ultrasonic wavelength (λ) over the bandwidth investigated. The sediment was boiled for 1 h to remove air bubbles and to saturate it with water. The water-saturated sediment was then poured into a small rectangular container immersed in water with one aperture at its top. The sediment container has the dimensions of 100 mm $\times$ 100 mm $\times$ 50 mm. The thickness of 50 mm was chosen to optimize the measurement of velocity dispersion and to avoid excessive attenuation of ultrasonic pulses. In order to minimize the transmission losses at the surfaces of the container perpendicular to wave propagation, its front and back walls (i.e., the faces which measured 100×100 mm^2) were made from thin plastic films 100 μm thick.
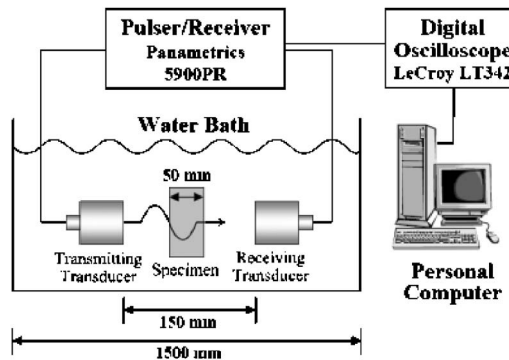
FIG. 1. Schematic diagram of the experimental setup for ultrasonic measurements in through-transmission geometry.

B. Ultrasonic measurements

Figure 1 illustrates the schematic diagram of the experimental setup for ultrasonic measurements in through-transmission geometry. A sediment specimen and transducers were immersed in a water bath with the dimensions of 650 mm $\times$ 750 mm $\times$ 1500 mm at room temperature. Two different matched pairs of unfocused, broadband transducers with a diameter (D) of 25.4 mm and center frequencies of 0.5 MHz (Panametrics V301) and 1.0 MHz (Panametrics V302) were used in order to cover the wide range of frequencies of interest. The opposing faces of coaxially aligned transducers were separated by a distance of 150 mm, greater than the near-field distances ($D^2/4\lambda$) of 53 and 108 mm for the 0.5 and the 1.0 MHz transducers. A 200 MHz pulser/receiver (Panametrics 5900PR) was used to generate pulses and to receive signals. Received signals were averaged over 100 pulses in time domain using a 500 MHz digital storage oscilloscope (LeCroy LT342) and stored on a computer for off-line analysis.

In order to measure the phase velocity, the received signals were recorded with and without the sediment specimen in the acoustic path. The frequency-dependent phase velocity $c_p(\omega)$ was determined by

$$c_p(\omega) = \frac{c_w}{1 - [c_w \Delta\phi(\omega)/\omega d]}, \tag{1}$$

where ω is the angular frequency of the wave, d is the thickness of the specimen, and $\Delta\phi(\omega)$ is the difference in unwrapped phases of the received signals with and without the specimen. The unwrapped phase difference $\Delta\phi(\omega)$ was calculated by taking the fast Fourier transform (FFT) of the digitized received signal, as follows (Wear, 2000). The phase of the signal at each frequency was taken to be the inverse tangent of the ratio of the imaginary to real part of the FFT at that frequency. Since the inverse tangent function yields principal values between $-\pi$ and π, the phase had to be unwrapped by adding an integer multiple of 2π to all frequencies above each frequency where a discontinuity appeared. The temperature-dependent speed of sound in distilled water, c_w, is given by (Kaye and Laby, 1995)

Lee *et al.*: Acoustic properties of a water-saturated sediment

$$c_w = 1402.9 + 4.835T - 0.047016T^2 + 0.00012725T^3, \tag{2}$$

where T is the temperature in °C. The estimates of Eq. (2) were consistent with the measured values in the water bath given the precision of the digital thermometer used. Measurements of phase velocity were also repeated with the specimen reversed. A total of 10 measurements (5 measurements in each direction) were obtained for a mean value, repositioning the specimen after each measurement. The sound speed in the sediment, ranging approximately from 1590 to 1680 m/s, is sufficiently close to that in fresh water at room temperature, 1483 m/s, that potential diffraction-related errors in this substitution method caused by the disparity in speeds between the two media may be ignored (Verhoef *et al.*, 1985; Xu and Kaufman, 1993).

The attenuation coefficient was determined by using the same signal acquired for phase velocity measurements, as follows. The FFT was taken to obtain the amplitude spectrum of the signal. The signal loss as a function of frequency was obtained by subtracting the logarithm of the amplitude spectrum of the signal through the sediment specimen from that through water only, and divided by the thickness of the specimen to determine the attenuation coefficient in units of dB/cm. Finally, the attenuation coefficient was corrected by taking into account the transmission losses at each interface, water/sediment and sediment/water, over the range of frequencies of interest. Ten measurements of attenuation coefficient in the specimen were averaged to obtain a mean value.

FIG. 2. (a) Temporal signals and (b) pressure spectra measured with and without the sediment in the acoustic path for the pair of 0.5 MHz transducers.

III. RESULTS AND DISCUSSION

A. Temporal signals and pressure spectra

Figures 2 and 3 show the temporal signals and their corresponding pressure spectra measured with and without the sediment in the acoustic path, using the two different pairs of transducers with the center frequencies of 0.5 and 1.0 MHz. The temporal signals observed in the sediment correspond to the fast longitudinal wave. In the present study, no attempt was made to measure the slow wave in the sediment. As shown in Figs. 2 and 3, each sediment signal exhibits an earlier arrival time than that observed in water. This is because the sound speed in the sediment is greater than that in water. It is notable that the sediment signal for the pair of 1.0 MHz transducers in Fig. 3(a) exhibits pulse elongation with the higher frequencies arriving later, characteristic of negative dispersion. As can be seen in spectra, the center frequencies of the pulses were clearly shifted to lower frequencies, resulting from increasing attenuation with frequency. This is more pronounced for the pulse centered at 1.0 MHz than that at 0.5 MHz. The usable frequency bandwidths were taken to be from 0.3 to 0.7 MHz for the pair of 0.5 MHz transducers and from 0.5 to 1.0 MHz for the pair of 1.0 MHz transducers.

B. Frequency dependence of phase velocity

The frequency-dependent phase velocity in the water-saturated sandy sediment was experimentally measured and theoretically predicted by the Biot and the GS models. The experimental results were plotted as a function of frequency in Fig. 4, over the usable frequency bandwidths of the transducers, i.e., from 0.3 to 0.7 MHz for the 0.5 MHz transducer measurements (circles) and from 0.5 to 1.0 MHz for the 1.0 MHz transducer measurements (asterisks). The error bars denote the standard deviations of ten measurements and represent the random uncertainties of the measurement procedure. It should be noted that there is some overlap of the two frequency bandwidths, i.e., from 0.5 to 0.7 MHz, where the phase velocities for the two sets of transducers agree within the experimental uncertainty. As clearly seen in Fig. 4, the average of measurements of phase velocity exhibits the significant negative dispersion, with the velocity ranging from 1680±5 m/s at 0.3 MHz to 1590±4 m/s at 1.0 MHz. The mean rate of decline in phase velocity was 120±20 m/s/MHz between 0.3 and 1.0 MHz. This negatively sloped trend is similar to that measured in water-saturated sandy sediments from 0.2 to 1.2 MHz by Moussatov *et al.* (1998) (1760–1680 m/s).

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Lee *et al.*: Acoustic properties of a water-saturated sediment   2555
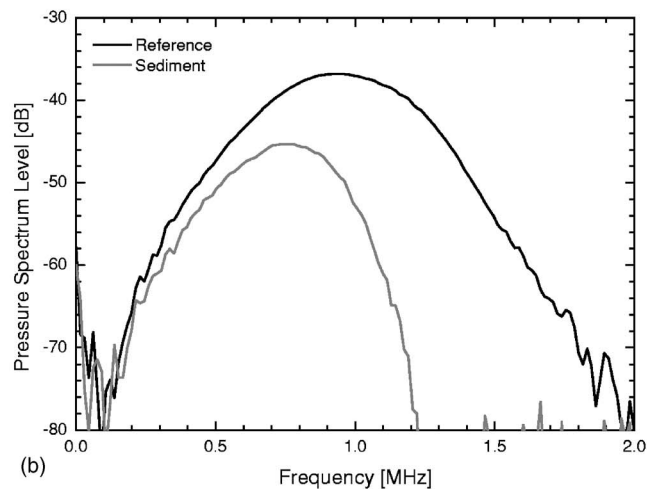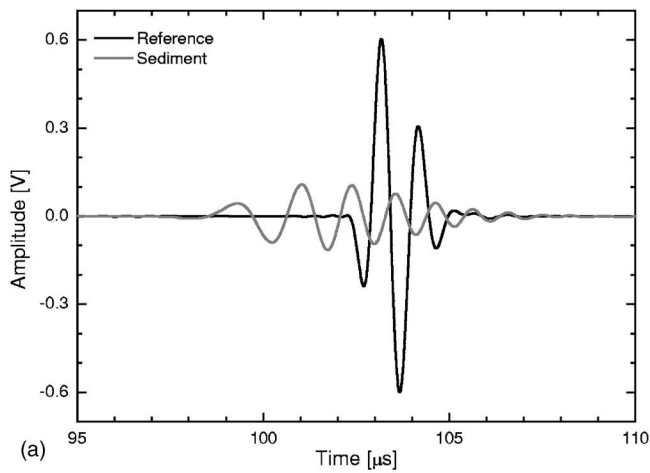
FIG. 3. (a) Temporal signals and (b) pressure spectra measured with and without the sediment in the acoustic path for the pair of 1.0 MHz transducers.

The solid and the dashed curves in Fig. 4 represent the phase velocities predicted by the Biot and the GS models, respectively. The input parameters of the Biot and the GS models for a water-saturated sandy sediment are summarized
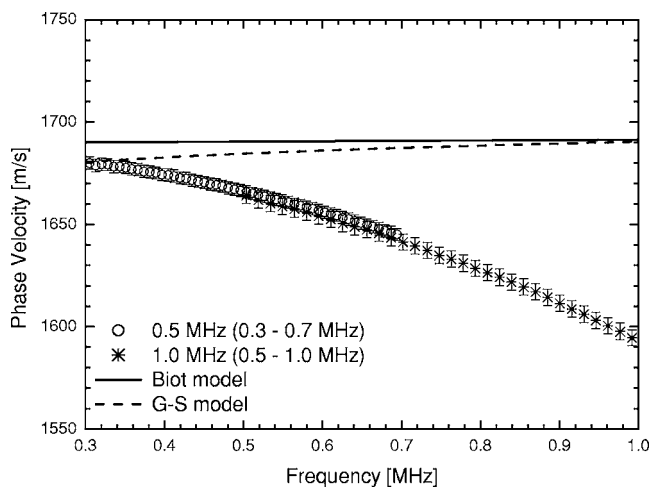
FIG. 4. Experimental and theoretical phase velocities plotted as a function of frequency for the fast longitudinal wave in the water-saturated sandy sediment.

TABLE I. Input parameters of the Biot model for a water-saturated sandy sediment.

Parameter	Value
Density of sand grain (ρ_s)	2650 kg/m^3
Density of fluid (ρ_f)	1000 kg/m^3
Bulk modulus of sand grain (K_s)	36.9 GPa
Bulk modulus of fluid (K_f)	2.2 GPa
Bulk modulus of frame (K_b)	66.3 MPa
Shear modulus of frame (μ_b)	24.9 MPa
Bulk log decrement (δ_k)	0.15
Shear log decrement (δ_μ)	0.2
Viscosity of fluid (η)	0.001 Pa s
Permeability (κ)	1×10^{-10} m^2
Pore size parameter (a)	58.2 μm
Tortuosity (α)	1.73
Porosity (β)	0.408

in Tables I and II. A large amount of information can be found in the literature about the determination of the parameters required for these two models (Chotiros, 1995; Buckingham, 1997, 2000, 2005; Williams *et al.*, 2002). The sediment depth parameter, $d=0.063$ m, in Table II was determined by curve fitting the prediction of the GS model to the experimental data of phase velocity at 0.3 MHz. As seen in Fig. 4, the Biot model predicts the approximately nondispersive phase velocity over the frequency range from 0.3 to 1.0 MHz and the GS model exhibits the slightly positive dispersion. This suggests that both of these models may not be suitable at very high frequencies for the dispersion of phase velocity in water-saturated sediments.

Multiple scattering models have shown to be useful in gaining an insight into wave propagation in unconsolidated granular composites saturated with fluid. Schwartz and Plona (1984) reported success in applying multiple scattering techniques (Nicholson and Schwartz, 1982; Tsang *et al.*, 1982) to the dispersion of phase velocity and attenuation in the glass- and the Plexiglas™-grain composites from 0.3 to 2.0 MHz. They used suspensions comprised of water-saturated, roughly monosized spherical grains with a porosity of 0.380. Especially, the glass grain had similar acoustic properties (

TABLE II. Input parameters of the GS model for a water-saturated sandy sediment.

Parameter	Value
Density of grain (ρ_s)	2650 kg/m^3
Density of fluid (ρ_f)	1000 kg/m^3
Bulk modulus of grain (K_s)	36.9 GPa
Bulk modulus of fluid (K_f)	2.2 GPa
Longitudinal coefficient (γ_{p0})	388.8 MPa
Shear coefficient (γ_{s0})	45.9 MPa
Strain-hardening index (n)	0.0851
rms grain roughness (Δ)	4.55 μm
Reference depth in sediment (d_0)	0.3 m
Reference grain diameter (u_0)	1000 μm
Reference porosity (β_0)	0.377
Depth in sediment (d)	0.063 m
Grain diameter (u_g)	425 μm
Porosity (β)	0.408
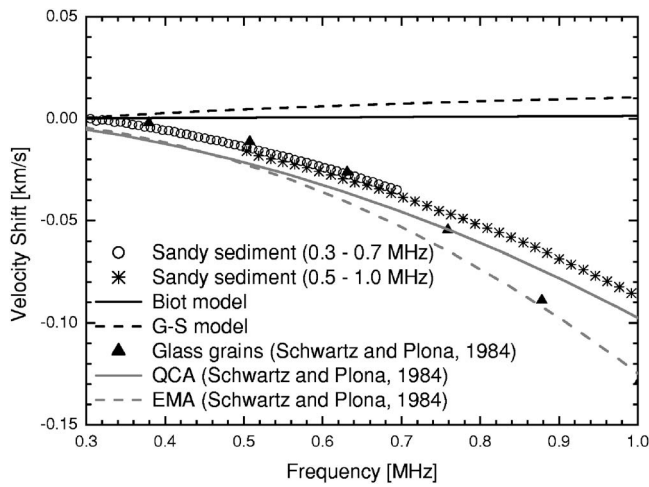
FIG. 5. Experimental and theoretical velocity shifts plotted as a function of frequency for the fast longitudinal wave in the water-saturated sandy sediment and in the glass-grain composite.

FIG. 6. Experimental and theoretical attenuation coefficients plotted as a function of frequency for the fast longitudinal wave in the water-saturated sandy sediment.

$\rho_s = 2480$ kg/m^3, $c_s = 5850$ m/s) and a mean diameter (545 μm) to those of the sand grain of the sediment used here. They compared the model predictions with the experimental data and concluded that, at high frequencies, the Plexiglas-grain composite exhibits the stronger scattering effects than the glass. It is notable that both unconsolidated composites showed the significant negative dispersion of phase velocity and the multiple scattering models, such as the familiar quasicrystalline approximation (QCA) and the self-consistent effective medium approximation (EMA), could successfully explain the negative dispersion (Schwartz and Plona, 1984). Figure 5 compares the predictions of these multiple scattering models for the glass-grain composite against the experimental data for the water-saturated sandy sediment measured here. It can be seen that the experimental results for the water-saturated sandy sediment presented in this study are also in general agreement with their predictions for the negative dispersion in the glass-grain composite. Future studies will apply the multiple scattering techniques to elucidate the characteristics of dispersion in the water-saturated sediment in more depth.

C. Frequency dependence of attenuation coefficient

Figure 6 shows the experimental and theoretical attenuation coefficients plotted as a function of frequency for the fast longitudinal wave in the water-saturated sandy sediment over the same frequency range as shown in Fig. 4. The error bars denote the standard deviations of ten measurements. It is shown that the experimental measurements are in good agreement with each other over the overlapped frequency bandwidth (from 0.5 to 0.7 MHz). The solid and the dashed curves in Fig. 6 represent the attenuation coefficients predicted by the Biot and the GS models, using the input parameters in Tables I and II, respectively. As seen in Fig. 6, the attenuation data increase nonlinearly with frequency whereas both the Biot and the GS models yield the attenuation coefficient that is almost linear in frequency from 0.3 to 1.0 MHz. It can also be seen that the Biot model consistently underestimates the experimental measurement of at-

tenuation. This is because the Biot model predicts absorption due to the viscous dissipation of the pore fluid only. In contrast, the prediction obtained from the GS model is larger than the measurement over most of the frequency range.
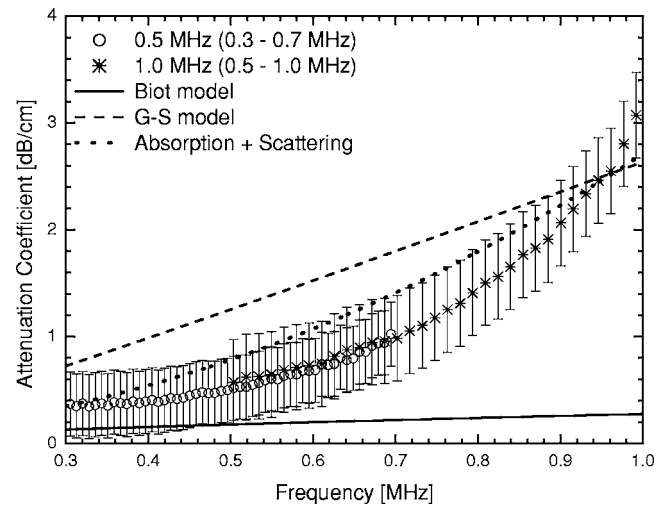
Ohkawa (2006) found an excellent agreement between the SAX99 attenuation data and the calculated attenuation coefficient given by the algebraic sum of absorption and scattering components. The absorption attenuation corresponds to the viscous dissipation of the pore fluid following the $f^{1/2}$ dependence, as the Biot model predicts, and the scattering attenuation is given by the empirical f^2 dependence obtained by curve fitting to the experimental data for water-saturated fine sands with a mean grain diameter of 230 μm measured by Seifert et al. (1999). The dotted curve in Fig. 6 represents the total attenuation coefficient, α, in units of dB/cm, calculated using

$$\alpha = \alpha_a + \alpha_s, \qquad (3)$$

where α_a is the absorption of the Biot model and α_s is the scattering obtained by curve fitting to the experimental data of attenuation coefficient for the water-saturated sandy sediment measured here, expressed as

$$\alpha_s = 0.28 \times 10^{-10} \times 20 \log(e) f^2, \qquad (4)$$

where f is the frequency of the wave in hertz. As found in Fig. 6, the total attenuation coefficient shows a reasonable agreement with the experimental data for overall frequencies. These findings underpin the fact that the scattering is the dominant cause of attenuation in the water-saturated sandy sediment with a mean grain diameter of 425 μm from 0.3 to 1.0 MHz. However, the scattering attenuation becomes weaker at lower frequencies and effects of the fluid viscosity may become more important. The role of absorption and scattering in attenuation at low and high frequencies provides an explanation for the difference observed when comparing in situ and laboratory measurements.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Lee et al.: Acoustic properties of a water-saturated sediment 2557

IV. CONCLUSIONS

We have investigated the frequency dependencies of phase velocity and attenuation coefficient for the fast longitudinal wave in a water-saturated sandy sediment over the frequency range from 0.3 to 1.0 MHz. The experimental data of phase velocity exhibited the significant negative dispersion, with the mean rate of decline of 120 ± 20 m/s/MHz. The Biot model predicted the approximately nondispersive phase velocity and the GS model exhibited the slightly positive dispersion. In contrast, the predictions of the multiple scattering models, such as the familiar QCA and the self-consistent EMA, for the negative dispersion in the glass-grain composite were in general agreement with the experimental data for the water-saturated sandy sediment measured here. The experimental data of attenuation coefficient was found to increase nonlinearly with frequency from 0.3 to 1.0 MHz. However, both the Biot and the GS models yielded the attenuation coefficient increasing almost linearly with frequency. The total attenuation coefficient given by the algebraic sum of absorption and scattering components showed a reasonable agreement with the experimental data for overall frequencies. This study suggests that the scattering is the principal mechanism responsible for the variations of phase velocity and attenuation coefficient with frequency in water-saturated sandy sediments at high frequencies.

ACKNOWLEDGMENTS

Biot, M. A. (**1956a**). "Theory of propagation of elastic waves in a fluid-saturated solid. I. Low-frequency range," J. Acoust. Soc. Am. **115**, 168–178.

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

Biot, M. A. (**1962**). "Generalized theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**, 1254–1264.

Buckingham, M. J. (**1997**). "Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments," J. Acoust. Soc. Am. **102**, 2579–2596.

Buckingham, M. J. (**2000**). "Wave propagation, stress relaxation, and grain-to-grain shearing in saturated, unconsolidated marine sediments," J. Acoust. Soc. Am. **108**, 2796–2815.

Buckingham, M. J. (**2005**). "Compressional and shear wave properties of marine sediments: Comparisons between theory and data," J. Acoust. Soc. Am. **117**, 137–152.

Chotiros, N. P. (**1995**). "Biot model of sound propagation in water-saturated sand," J. Acoust. Soc. Am. **97**, 199–214.

Chotiros, N. P., and Isakson, M. J. (**2004**). "A broadband model of sandy ocean sediments: Biot-Stoll with contact squirt flow and shear drag," J. Acoust. Soc. Am. **116**, 2011–2022.

Hamilton, E. L. (**1987**). "Acoustic properties of sediments," in *Acoustics and Ocean Bottom*, edited by A. Lara-Saenz, C. Ranz-Guerra, and C. Carbo-Fite (Consejo Superior de Investigacions Cientificas, Madrid, Spain), pp. 3–58.

Hampton, L. D. (**1967**). "Acoustic properties of sediments," J. Acoust. Soc. Am. **42**, 882–890.

Kaye, G. W. C., and Laby, T. H. (**1995**). *Tables of Physical and Chemical Constants and Some Mathematical Functions* (Longman, London).

Moussatov, A., Guillon, L., Ayrault, C., and Castagnede, B. (**1998**). "Experimental study of the dispersion of ultrasonic waves in sandy sediments," C. R. Acad. Sci., Ser. IIb Mec. Phys. Astron. **326**, 433–439.

Nicholson, D., and Schwartz, L. (**1982**). "Structure-induced minimum in the average spectrum of a liquid or amorphous metal," Phys. Rev. Lett. **49**, 1050–1053.

Nolle, A. W., Hoyer, W. A., Mifsud, J. F., Runyan, W. R., and Ward, M. B. (**1963**). "Acoustical properties of water-filled sands," J. Acoust. Soc. Am. **35**, 1394–1408.

Ohkawa, K. (**2006**). "Confirmation of the Biot theory for water-saturated sands at high frequencies and effects of scattering on the attenuation of sound waves (L)," J. Acoust. Soc. Am. **119**, 709–711.

Schwartz, L., and Plona, T. J. (**1984**). "Ultrasonic propagation in close-packed disordered suspensions," J. Appl. Phys. **55**, 3971–3977.

Seifert, P. K., Kaelin, B., and Johnson, L. R. (**1999**). "Effect on ultrasonic signals of viscous pore fluids in unconsolidated sand," J. Acoust. Soc. Am. **106**, 3089–3094.

Stoll, R. D. (**2002**). "Velocity dispersion in water-saturated granular sediment," J. Acoust. Soc. Am. **111**, 785–793.

Stoll, R. D., and Bryan, G. M. (**1970**). "Wave attenuation in saturated sediments," J. Acoust. Soc. Am. **47**, 1440–1447.

Tang, D., Briggs, K. B., Williams, K. L., Jackson, D. R., and Thorsos, E. I. (**2002**). "Fine-scale volume heterogeneity measurements in sand," IEEE J. Ocean. Eng. **27**, 546–560.

Tsang, L., Kong, J. A., and Habashy, T. (**1982**). "Multiple scattering of acoustic waves by random distribution of discrete spherical scatterers with the quasicrystalline and Percus-Yevick approximation," J. Acoust. Soc. Am. **71**, 552–558.

Turgut, A., and Yamamoto, T. (**1990**). "Measurements of acoustic wave velocities and attenuation in marine sediments," J. Acoust. Soc. Am. **87**, 2376–2383.

Verhoef, W. A., Cloostermans, M. J. T. M., and Thijssen, J. M. (**1985**). "Diffraction and dispersion effects on the estimation of ultrasound attenuation and velocity in biological tissues," IEEE Trans. Biomed. Eng. **BME-32**, 521–529.

Wear, K. A. (**2000**). "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.

Williams, K. L., Jackson, D. R., Thorsos, E. I., Tang, D., and Schock, S. G. (**2002**). "Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous media," IEEE J. Ocean. Eng. **27**, 413–428.

Wingham, D. J. (**1985**). "The dispersion of sound in sediment," J. Acoust. Soc. Am. **78**, 1757–1760.

Xu, W., and Kaufman, J. J. (**1993**). "Diffraction correction methods for insertion ultrasound attenuation estimation," IEEE Trans. Biomed. Eng. **40**, 563–570.

Diversity in ambient noise in European freshwater habitats: Noise levels, spectral profiles, and impact on fishes

Lidia Eva Wysocki,[a] Sonja Amoser,[b] and Friedrich Ladich[c]
Department of Behavioural Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

The detectability of acoustic signals depends on the hearing abilities of receivers and the prevailing ambient noise in a given habitat. Ambient noise is inherent in all terrestrial and aquatic habitats and has the potential to severely mask relevant acoustic signals. In order to assess the detectability of sounds to fishes, the linear equivalent sound pressure levels (L_{Leq}) of twelve European freshwater habitats were measured and spectra of the ambient noise recordings analyzed. Stagnant habitats such as lakes and backwaters are quiet, with noise levels below 100 dB re 1 μPa (L_{Leq}) under no-wind conditions. Typically, most environmental noise is concentrated in the lower frequency range below 500 Hz. Noise levels in fast-flowing waters were typically above 110 dB and peaked at 135 dB (Danube River in a free-flowing area). Contrary to stagnant habitats, high amounts of sound energy were present in the high frequency range above 1 kHz, leaving a low-energy "noise window" below 1 kHz. Comparisons between the habitat noise types presented here and prior data on auditory masking indicate that fishes with enhanced hearing abilities are only moderately masked in stagnant, quiet habitats, whereas they would be considerably masked in fast-flowing habitats. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713661]

I. INTRODUCTION

In the aquatic environment, sound is a key signal carrier because it is not attenuated as fast as light or chemical substances, and can be propagated over long distances (Hawkins and Myrberg, 1983; Bradbury and Vehrencamp, 1998; Mann, 2006). This makes sound an important carrier of information for fish and for other aquatic vertebrates. Exploitation of biotic and abiotic acoustic cues may be vital for detecting predators, prey, mates, and competitors but also for acoustic orientation, i.e., detecting coastlines, torrents, wind, and water currents.

The detectability of relevant signals will be strongly affected not only by the hearing abilities of the receiver but also by the environment. Physical properties of the environment such as depth, bottom structure, vegetation, and temperature and salinity clines lead to surface and bottom reflections, causing complex multiple sound propagation paths and therefore often considerable signal distortion and limitation of the propagation distance (Gerald, 1971; Fine and Lenhardt, 1983; Crawford *et al.*, 1997; Mann and Lobel, 1997; Ladich and Bass, 2003).

Another limitation in signal detectability is the presence of an ambient background noise floor inherent in every habitat—aquatic or terrestrial. Such noise emanates from biotic (sounds from animals) and abiotic (i.e., wind, precipitation, surf) sources. Urick (1983) defined underwater ambient noise in a strict sense as the residual noise background in the absence of individual identifiable sources that may be considered the natural noise environment. However, it is often difficult to identify specific noise sources due to a lack of knowledge about the biotic or abiotic noise components. We therefore apply the term "habitat or ambient noise" throughout this paper. Habitat in an ecological sense is defined as the place where a particular species lives and grows, i.e., the environment that influences a species' population. Hence, the broad term habitat or ambient noise covers every surrounding noise floor with which an animal is confronted in its natural environment. This noise is likely to vary in level, as well as in spectral composition, between different aquatic ecosystems depending on abiotic and biotic factors; originally, it did not include clearly recognizable anthropogenic sources such as noise emanating from shipping.

This ambient noise can impair the detection of a relevant sound, a phenomenon termed masking (Fletcher, 1940). The masking effect, i.e., the amount by which a signal must be louder in order to still be detected in the presence of background noise of a given level, has been demonstrated to largely depend on the overall hearing sensitivity of a particular fish species (Tavolga, 1967; Champman, 1973; Wysocki and Ladich, 2005). Recently, the masking effect of habitat noise has been studied in several freshwater species (Amoser and Ladich, 2005; Scholz and Ladich, 2006). Amoser and Ladich (2005) showed that hearing-sensitive fish such as carp are only moderately masked by quiet habitat noise levels of standing waters, whereas they are highly affected by stream and river noise, in whose presence acoustic signals must be up to 50 dB louder in their best hearing range (0.5–1.0 kHz) than in the quiet habitats to be detectable. In contrast, the hearing thresholds of perch (a species with much poorer hearing sensitivity than carp) were only slightly affected by the highest habitat noise levels presented. Scholz and Ladich (2006) demonstrated that the topmouth minnow (*Pseudorasbora parva*) can easily intercept feeding sounds

[a] Electronic mail: lidia.wysocki@univie.ac.at
[b] Electronic mail: sonja.amoser@univie.ac.at
[c] Electronic mail: friedrich.ladich@univie.ac.at
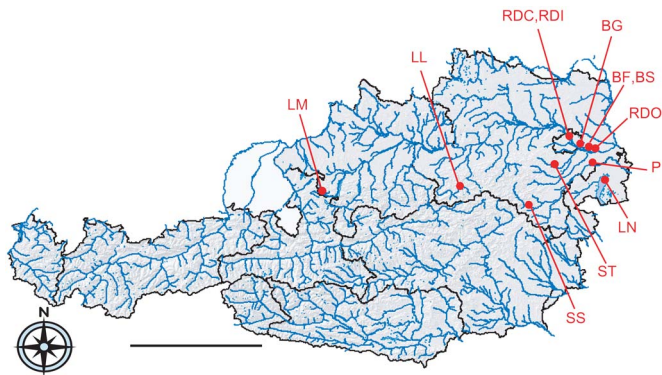
FIG. 1. (Color online) Overview of the study sites. In alphabetical order: BF: Backwater Fadenbach; BG: Backwater Gänsehaufen Traverse; BS: Backwater Schönauer Traverse; LL: Lake Lunz; LM: Lake Mondsee; LN: Lake Neusiedl; P: pond; RDC: Danube Channel; RDI: Danube River at Danube Island; RDO: Danube River at Orth; SS: Schwarza stream; and ST: Triesting stream. Bar indicates 120 km. Map of Austria from the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management© BMLFUW 1998-2004.

TABLE I. Geomorphologic information about the study sites. In alphabetical order: BF: Backwater Fadenbach, BG: backwater Gänsehaufen Traverse, BS: Backwater Schönauer Traverse, LL: Lake Lunz; LM: Lake Mondsee; LN: Lake Neusiedl; P: pond; RDC: Danube Channel; RDI: Danube River at Danube Island, RDO: Danube River at Orth; SS: Schwarza stream; ST: Triesting stream; and SL: sea level.

Study site	Geographical position	Altitude (m above SL)	Total depth (m)	Hydrophone depth (m)	Recording date
LM	47.9°N, 13.4°E	481	68	5	3 July
LL	47.9°N, 15.1°E	608	34	1.7	5 June
LN	47.8°N, 16.8°E	115	2	1	4 June
BF	48.1°N, 16.7°E	149	<0.5	0.5	7 July
BS	48.1°N, 16.6°E	153	<1	0.5	4 May
BG	48.2°N, 16.6°E	149	<1	0.4	4 May
P	48.1°N, 17.0°E	163	1.5	0.8	3 April
SS	47.7°N, 15.9°E	457	<0.5	0.4	3 July
ST	48.0°N, 16.1°E	346	<0.5	0.3	4 April
RDC	48.3°N, 16.4°E	149	2	0.5	4 June
RDI	48.3°N, 16.4°E	146	3	0.5	4 June
RDO	48.1°N, 16.7°E	149	2	0.6	4 July

of conspecifics over distances of several dozen centimeters in one of its typical stagnant habitats (a shallow lake), an ability which would be considerably impaired in louder habitats such as creeks or streams.

Therefore, every speculation on acoustic communication ranges in the field as well as on the detectability of various noise sources must consider the potential masking effects of prevailing ambient noise. While several studies have characterized habitat noise in the marine environment (e.g. Wenz, 1962; Cato, 1976; Urick, 1983; McConnell *et al.*, 1992; Samuel *et al.*, 2005; McDonald *et al.*, 2006), including reefs (Tolimieri *et al.*, 2004; Egner and Mann, 2005), only sparse data are available for freshwater habitats (Bom, 1969; Boussard, 1981; Lugli and Fine, 2003; Lugli *et al.*, 2003; Amoser *et al.*, 2004); especially comparative approaches are lacking.

The present study therefore comparatively describes the sound pressure levels and the spectral composition of ambient noise types in various central European freshwater habitats. Detailed knowledge about the ambient noise in freshwater habitats and about their fish community composition, together with a comparison of the ambient noise types whose effects on fish hearing have already been investigated (Amoser and Ladich, 2005; Scholz and Ladich, 2006), will allow the amount of masking for various fish species to be assessed. This, in turn, will highlight the limitations of fish acoustic orientation and communication.

II. MATERIAL AND METHODS

A. Study sites

Twelve different study sites throughout Austria, which represent the variety of freshwater habitats in Central Europe, were chosen. These comprised stagnant or slowly flowing freshwater habitats (lakes and river backwaters) as well as fast-flowing freshwaters (small streamlets and the large Danube River). Figure 1 provides an overview on the location of the study sites within Austria, and Table I lists further information about the sites.

Three different lakes were chosen in order to account for the variety of lakes encountered in Austria. As can be seen in Table I these lakes differ not only in the area they cover [Lake Lunz (LL): 0.7 km²; Lake Mondsee (LM): 14 km²; and Lake Neusiedl [LN]: 321 km²] but also in the total water depth and the altitude.

The backwaters of the Danube River [Fadenbach (BF), Schönauer Traverse (BS), and Gänsehaufen Traverse (BG)] are all part of the Danube Floodplain National Park. The water depth is approximately 0.5 m, but can exceed 1 m during high waters.

The pond (P) is located in Prellenkirchen southeast of Vienna, measures 32×22 m with a depth of about 1.8 m and is populated by cyprinids.

The Schwarza (SS) and the Triesting stream (ST) are typical creeks of the Alpine foothills. The bed of the Schwarza contains bedrock (according to the substrate type classification of Yamazaki *et al.*, 2006), making its flow more intermittent, with areas of retained waters (our randomly selected study site was in one such area) and areas where the water can flow freely and fast. Contrary to the Schwarza, the substrate type in the Triesting stream is cobble (65–256 mm) to boulders (>256 mm). At our study site we measured a mean current velocity of 0.5 m/s.

The Austrian part of the Danube River is a mountainous river with a mean slope of 40 cm km⁻¹ and a current velocity of 1–3 m s⁻¹. Its water regime is mainly dominated by the snowmelt in the high Alps; therefore, high water occurs usually from late spring to summer, although precipitation-caused floodings can occur throughout the year. Three different study sites were chosen: one in a free-flowing part [Danube River at Orth (RDO)], the second in a retained part [Danube River at Danube Island (RDI)] of the main river. The last of these study sites was in a now regulated former side-arm of the Danube River flowing through the center of Vienna [Danube Channel (RDC)]. The latter two habitats are affected by human activities (ship and road traffic) but were

included into this study because they are populated by numerous fish species and the fish there have to deal with these modifications of their natural environment.

All recordings were made in spring and early summer of 2003–2005 between 10 a.m. and 2 p.m. (Table I), the main reproductive season of most fish species. The aim of this study was to give a first insight into the variability of ambient noise levels and spectra in various freshwater habitats. We decided to measure at just one site per habitat and concentrate on a wide spectrum of habitats, although by doing this the potential variation within each habitat might be underestimated. However, dealing with this variability would have gone beyond the scope of this study. Three noise types already described in Amoser and Ladich (2005) (the noises of the Danube River at Orth, Triesting stream, and Lake Neusiedl) were included in the present study for comparative purposes; they allow us to estimate the effects of noise in the other habitats on fishes.

B. Noise measurements and recording

Underwater noise was recorded on a DAT recorder (Sony TCD 100) using a hydrophone (Brüel & Kjaer 8101) powered by a power supply (Brüel & Kjaer 2804). Due to the different water depths of the habitats (Table I) it was not possible to position the hydrophone always in the same depth. We managed to place the hydrophone in a depth of 0.3–1 m most of the time, except for the measurements in the lakes (Table I). The minimum recording period was 1 min, but for most recordings this period was expanded to 3 min. Before and after each recording the sound pressure level (SPL) of the ambient noise was measured for a period of 1 min using a sound level meter (Brüel & Kjaer 2238 Mediator), which was connected to one output of the power supply. Thus, it was possible to attribute SPL values to each of the recordings. Two SPL measures were obtained: (1) the instantaneous SPL (L_{LSP}, L-weighted, 5 Hz to 20 kHz, rms fast), to assess the variability of habitat noise levels over time, and (2) the equivalent continuous SPL (L_{Leq}), averaged over 60 s. The L_{Leq} is a measure of the averaged energy in a varying sound field and is commonly used to assess environmental noise (ISO 1996, 2003). During each measurement the L_{LSP} was noted every 5 s, and at the end of the 1 min period we noted the L_{Leq} as well (the B & K Mediator allows parallel reading of four different noise measures), except for Lake Mondsee and the Schwarza stream, where only one L_{Leq} measure was noted before and after, respectively.

C. Sound analysis

For this comparative study, one sound recording of each habitat was chosen. It had to meet certain criteria like time of recording in spring and early summer (April to July, Table I), no anthropogenic noise (i.e., boat noise, road traffic noise) present with the exception of the two habitats, Danube Island and Danube Channel, where no such recordings could be obtained, and good weather conditions (i.e., no rain, no wind, no wind-related waves). The selected sound samples were digitized at a sampling rate of 44.1 kHz, and 60 s samples were analyzed using STX 3.7.4 developed by the Acoustic

Research Institute at the Austrian Academy of Sciences.

First relative noise spectra of the 60 s samples were calculated by fast Fourier transformations using a filter bandwidth of 1 Hz. Absolute spectra were calculated using the L_{Leq} measured immediately before/after the recordings following the algorithm described in Amoser *et al.* (2004) and Wysocki and Ladich (2005).

D. Statistical analysis

L_{LSP} levels between the different study sites were compared using a Kruskal-Wallis test because data were neither normally distributed nor were variances homogeneous. The Mann-Whitney-U-test was used for further pairwise comparisons. Furthermore, the overall difference between stagnant and flowing freshwaters was compared with a T-test using the L_{Leq} measured before and after each noise recording of each habitat noise type. All statistical tests except cross correlations were run using SPSS 12.0

In order to assess similarities in noise spectra between habitats cross correlations from 60 s noise segments of each habitat were calculated using STX 3.7.4. A cross correlation value of 1 indicates a strong correlation, with a plus indicating a positive correlation, i.e., the values of one curve tend to become larger when the values of the other curve become larger, too, and vice versa for negative values. We considered values from 1 to 0.9 as strongly correlated, values from 0.9 to 0.6 as correlated, and values from 0.6 to 0 as weakly correlated or not (Townend, 2002). Furthermore, to assess the effect of habitat noise types on the hearing of fishes, spectral profiles of habitats were compared within their hearing range (20–5000 Hz; Ladich and Popper, 2004) to those habitat noise types whose effects on fish hearing have been described previously (Amoser and Ladich, 2005). Therefore, cross correlations between the newly described noise types and noise types previously described in Amoser and Ladich (2005) were calculated. Since cross correlation coefficients are simply a measure for similarity in shape, i.e., in spectral composition without the absolute noise levels taken into account (Wood *et al.*, 2005), this information needs to be combined with overall absolute sound pressure levels in order to reliably assess potential masking effects of the noise.

III. RESULTS

A. Absolute noise levels

Instantaneous SPLs (L_{LSP}) differed significantly between the various study sites (Kruskall-Wallis: $\chi^2_{11,252}=238.6$, $p \leq 0.001$) (Fig. 2). Subsequent Mann-Whitney-U tests revealed that L_{LSPs} of Lake Mondsee and the backwater Schönauer Traverse [$U=256.0$, $p=0.495$] [Fig. 2(a)], the backwater Gänsehaufen Traverse and the pond [$U=59$, $p=0.436$, Fig. 2(b)] as well as Danube Island and Schwarza [$U=237.5$, $p=0.294$, Fig. 2(c)] did not differ significantly from each other, whereas the levels of all other site combinations were significantly different. The variability within the same study site was lowest at the site with the highest average L_{LSP} values: The difference between maximum and minimum levels measured was only 1 dB at Danube River at Orth (Table II), whereas it was 15 dB at Danube Island and
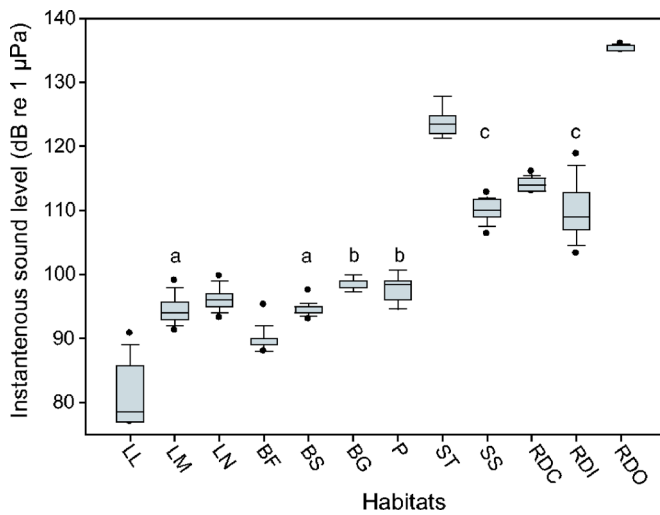
FIG. 2. (Color online) L_{LSP} values (median and quartiles) of the 12 different habitats. Indices indicating values that are not significantly different from each other (Mann-Whitney-U test).

12 dB at Lake Lunz, the site with lowest average SPLs (Table II). However, the average difference between maximum and minimum levels measured at the stagnant sites was the same as within fast-flowing waters (6.6 dB).

Continuous SPLs (L_{Leq}) ranged from 79 dB (Lake Lunz) to 135 dB (Danube River) (Table II). Stagnant waters had a L_{Leq} below 100 dB re 1 μPa (mean 92.8±1.85 SE) and differed significantly from flowing waters such as creeks and river sites, which were above 110 dB (mean 118.8±3.10 SE) (T-test, $t=-7.59$, df=21, $p<0.001$). The L_{Leq} values varied to some degree within the same habitat types, e.g., the levels in the seven stagnant habitats (lakes, pond, backwaters) differed by up to 20 dB, the three different study sites at the Danube River (Danube River at Orth, and Danube Channel, and Danube River at Danube Island) differed even more (up to 25 dB), and even within the two creeks (Triesting, Schwarza) there was more than 10 dB variance. Nonetheless, two groups of habitats are apparent: The more quiet stagnant and the noisier fast-flowing habitats.

TABLE II. Continuous (L_{Leq}) and instantaneous (L_{LSP}) noise levels encountered in the 12 freshwater habitats. All values are given in dB re 1 μPa. Before: Before noise recording, After: After noise recording.

	L_{Leq}			L_{LSP}	
Study site	Before	After	L_{LSP} (x±SE)	min	max
LL	80.5	78.6	81.0 ± 0.98	77	91
LM	94.0	a	94.3 ± 0.40	91	99
LN	97.7	95.7	96.0 ± 0.37	93	100
BF	90.5	88.2	89.5 ± 0.36	88	96
BS	94.7	94.0	94.4 ± 0.21	93	98
BG	98.9	97.8	98.7 ± 0.25	97	100
P	97.1	98.6	98 ± 1.18	96	100
SS	a	110.4	110.2 ± 0.42	108	113
ST	122.3	123.2	123.6 ± 0.60	121	129
RDC	114.3	114.2	114.3 ± 0.19	113	116
RDI	110.5	113.4	109.6 ±0.85	103	119
RDO	135.1	135.3	135.3 ± 0.09	135	136

aNot measured.

B. Spectral profiles

Amplitude spectra of habitat noises differed widely in composition (Fig. 3). All spectra revealed a decline in energy from 20 Hz down to 20 kHz, with a major decline at lowest frequencies. This decline could be very smooth and continuous such as in Lake Lunz or more irregular because of energy peaks such as in the Triesting stream and the Gänsehaufen backwater.

Lakes, backwaters, and the pond showed similar spectral profiles with main energies in the low frequency region followed by a rather steep decline in energy until 100 (backwaters) to 800 Hz (pond), followed by either a continuous, slow decline or a leveling off (BF, P), except for two backwaters which had a region of higher energy between 5.3 and 9.3 kHz (BG), or between 7 and 12 kHz (BS).

The running waters on the other hand showed a different frequency distribution pattern. Again main energies were found in the low frequencies, followed by a rapid decline to a minimum. In contrast to the standing waters, spectral levels then increased again, forming an area of lower spectral levels ("noise window") ranging from 170 to 450 Hz (in the Schwarza stream this window even ranged up to 2 kHz). This noise window was sharpest in the spectra of the Danube River at Orth and Danube River at Danube Island, both being in the main river. In the Danube Channel and the Triesting stream, the energy content was more fluctuating but the presence of such a noise window remained obvious.

In general, the energy in the 200 Hz to 5 kHz range was much higher in the streams and rivers (with the exception of SS) than in the stagnant water bodies.

C. Cross correlations

To assess the masking effect of the ambient noise types, the amplitude spectra of the different habitat noises were cross correlated with each other and with noise types whose effect is already known from Amoser and Ladich (2005): Lake Neusiedl, a backwater of the Danube River, Danube River at Orth and lab noise (Lab) (Table III). The spectral composition of stagnant habitats seems to be more similar in shape than those of fast-flowing habitats with cross correlation coefficients ranging from 0.58 to 0.86 (lakes) and from 0.80 to 0.93 (backwaters). The pond also showed rather high cross correlation coefficients with all other backwaters and the lab noise (Table III). The amplitude spectrum of the quietest habitat, Lake Lunz, was also strongly correlated to the lab noise of Amoser and Ladich (2005) ($r=0.96$, Table III). In contrast, the correlation coefficients could be very diverse between running waters. Whereas the cross correlation coefficients between the Danube River at Orth and all the other fast-flowing habitats were very low (ranging from −0.67 to 0.13, Table III), the spectrum of the Triesting was highly correlated to that of the site Danube Island ($r=0.86$) but not at all correlated to the second creek, the Schwarza ($r=-0.02$).
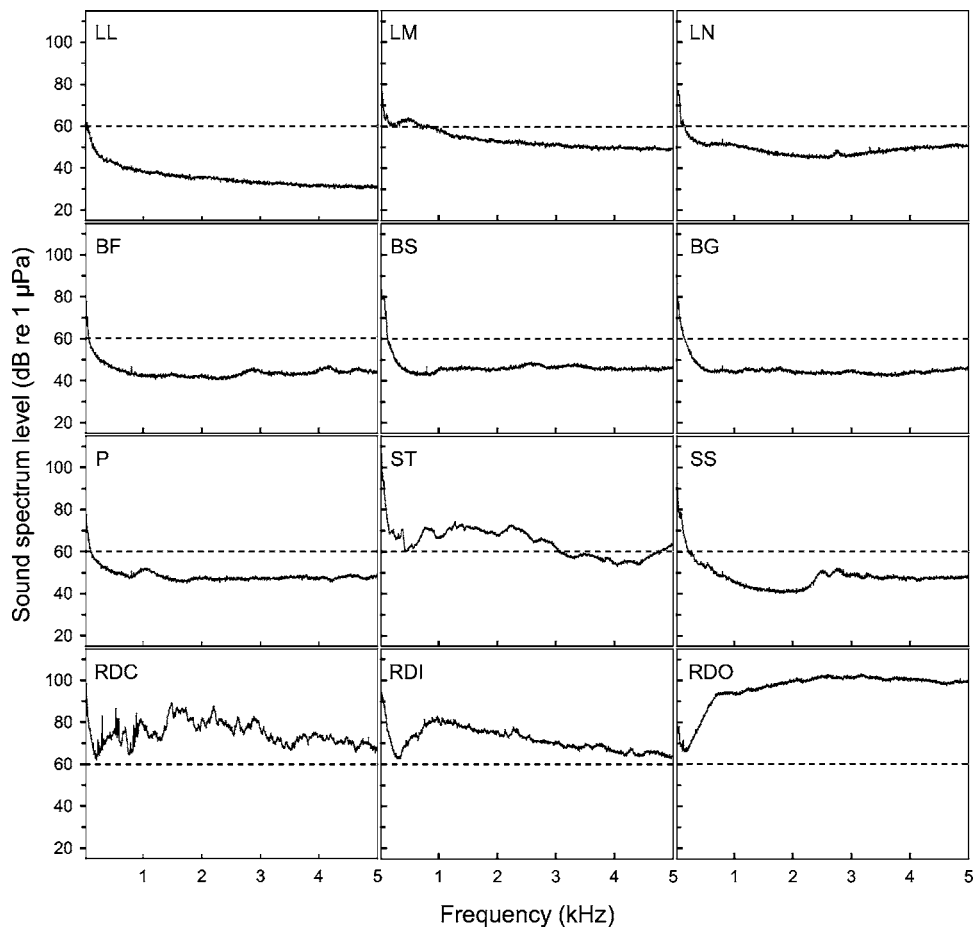
FIG. 3. Amplitude spectra of the 12 habitats. For abbreviations see Fig. 1. Frequency range 20 Hz to 5 kHz, bandwidth 1 Hz, overlap 50%, window: Blackman-Harris.

IV. DISCUSSION

A. Habitat noise levels

In the past decades, interest in the underwater ambient noise in different habitats (mainly in the marine environment) has grown. The diverse studies have applied different methods and measures of noise levels. Therefore, in order to avoid confusion, a short overview of the different values employed is given. The term "overall noise level" usually refers to broad band (i.e., covering the whole frequency range) root mean square SPLs—rms SPL values— (Richardson *et al.*, 1995). These levels are higher than the spectrum levels, which only render the energy content at particular frequencies or frequency bands. When describing the spectral composition, the given noise levels were either measured in frequency bands of the sound (e.g., 1/3 octave, octave) and expressed as pressure density spectra (unit: dB re

TABLE III. Cross correlation coefficients between the different habitat noises and combined with the corresponding masking noise reported in Amoser and Ladich (2005). For abbreviations see Fig. 1. BW=Backwater noise (in this case not identical to any of the backwater noises of this study). RDO, ST, LN [in italics because the identical noise was used in Amoser and Ladich (2005) and this study], BW and Lab spectra are from Amoser and Ladich (2005). Frequency range: 20 Hz–5 kHz, bandwidth: 1 Hz, 50% overlap, Blackman window. Bold values indicate strongly correlated spectra according to Townend (2002).

	LL	LM	*LN*	BF	BS	BG	P	SS	*ST*	RDC	RDI	*RDO*
LL		**0.86**	0.73	0.68	0.58	0.63	0.82	0.64	0.58	0.19	0.59	−0.86
LM			0.58	0.37	0.19	0.27	0.64	0.39	0.55	0.17	0.63	−0.72
LN				0.82	0.69	0.77	**0.91**	0.76	0.21	−0.11	0.27	−0.68
BF					0.83	**0.89**	0.82	**0.85**	0.20	0.07	0.12	−0.61
BS						**0.93**	0.74	0.75	0.28	0.16	0.26	−0.40
BG							0.80	0.80	0.24	0.09	0.21	−0.50
P								0.78	0.32	0.01	0.40	−0.76
SS									−0.02	−0.20	−0.01	−0.67
ST										0.63	**0.86**	−0.24
RDC											0.53	0.13
RDI												−0.22
BW[a]	**0.86**	0.65	**0.89**	**0.88**	0.80	0.84	**0.91**	0.83	0.39	0.09	0.39	−0.74
Lab[a]	**0.96**	**0.88**	0.75	0.67	0.52	0.59	0.83	0.66	0.50	0.12	0.52	−0.89

[a]Cross correlation with noise spectra from Amoser and Ladich (2005).

1 μPa^2/Hz), or given without a correction for filter band-width, so-called power spectra (unit: dB re 1 μPa). Unfortunately, very often this information is lacking, complicating direct comparisons between different studies.

The overall broadband continuous equivalent habitat noise level in the current study was lowest in Lake Lunz (79 dB L_{Leq}). This Alpine lake is fed and drained by a small creek with slow current and the water exchange rate is very slow. The creek flowing through the lake ("Seebach") did not contribute a significant amount of background noise in our recording, and its effect is probably more local. In contrast, the highest noise levels were found in the Danube River at Orth (135.3 dB L_{Leq}, Table II), mainly due to the high current velocities (1–3 m s^{-1}) as well as to the transportation of a larger water body and of cobble or gravel in the river bed. In between these two extremes, there was a great variability of overall noise levels. Nonetheless, the study sites can be separated into two large groups based on their overall broadband noise levels: stagnant habitats with L_{Leq} levels generally below 100 dB, and fast-flowing habitats with L_{Leq} levels above 110 dB.

Interestingly, when examining the variability of instantaneous noise levels (L_{LSP}), the situation is inverse: The noise levels at the noisiest study site (Danube River at Orth) were highly conservative (only a 1 dB variation, Table II), whereas variability in quieter habitats such as Danube Island and Lake Lunz were relatively high (15 and 12 dB, respectively). One possible explanation is that noise levels in the former are at such high levels that additional noise sources barely contribute to the overall level. On the other hand, in quieter habitats even slight changes in the noise environment (e.g., swimming activities of fishes and birds, insect sounds, but also noises of anthropogenic origin) are clearly detectable and thus contribute to the overall noise level.

Whereas data are available for noise levels in the ocean, data on freshwater systems—with which our measurements could be compared—are very sparse and often limited to spectral levels. While noise spectrum levels in the ocean are generally above 100 dB re 1 μPa (Wenz, 1962), the few freshwater data point to spectrum levels mostly below 100 dB re 1 μPa and of a higher variability. Exceptions have been reported. Hawkins and Johnstone (1978) mentioned that noise spectrum levels (1 Hz bands) in the River Dee are 5–10 dB higher than in the sea (Loch Torridon, Scotland). Data on noise levels in different European and one African river showed a high variability: Boussard (1981) measured ambient noise band levels (1/3-octave bands) ranging from 80–100 dB (re 1 μPa) in the River Meuse. Crawford *et al.* (1997) reported spectrum levels of up to 70 dB (1 Hz bands) at low frequencies, with a steady fall-off to about 40 dB above 1 kHz in the background noise of a floodplain of the Niger River (Mali, Africa). Lugli and Fine (2003) observed spectral levels (1 Hz bands) ranging from 70 dB in quiet places to 100 dB in fast-flowing areas. This is very similar to our data on the Danube River: depending on whether measurements were conducted in a retained (Danube River at "Danube Island") or a free-flowing area (Danube River at Orth), the noise spectrum levels ranged from about 60 to 100 dB (Fig. 2).

Data about noise levels in freshwater lakes, again, are very sparse. In an Italian Lake (Lake Sarzana), Bom (1969) encountered noise spectrum levels of 40–50 dB (octave bands). Nystuen (1986) measured spectral levels ranging from 40–60 dB (1 Hz bands) in Lake Clinton, IL. Amoser *et al.* (2004) observed higher spectral noise levels (1 Hz bands) in an Austrian Alpine lake (Lake Traunsee): about 80 dB at low frequencies and about 60 dB above 5 kHz. This is likely due to the Traun stream, which flows through that lake and contributes to these high levels. Our current data from three other lakes suggest that spectrum levels are below 65 dB at frequencies above 100 Hz and below 50 dB (or even below 35 dB in the case of Lake Lunz) above 5 kHz (Fig. 2).

In general, noise levels in freshwater habitats apparently depend primarily on the hydrology (abiotic sources), most notably the volume and speed of the water flow with cavitations and sediment transportation, whereas biotic sources (such as water insects) can only significantly contribute to the overall noise levels in stagnant or slowly flowing habitats with otherwise low noise levels.

B. Habitat noise spectra

Knowledge of sound spectra composition and the noise sources in specific frequency regions is also better explored in the marine environment. Wenz (1962) and Urick (1983) reported that the ambient noise in the ocean contains high amounts of energy in the low frequency regions (up to several hundred hertz), followed by a rapid decline with increasing frequency. Different noise sources contributed in the former study to different parts of the noise spectrum, e.g., oceanic traffic noise (i.e., the noise from distant shipping) seems to be the characteristic source for 10 Hz to 1 kHz, whereas wind-dependent noise (from spray and bubbles) contributes more to higher frequencies (100 Hz to 10 kHz). Biological noise has been observed at all frequencies from as low as 10 Hz to as high as 10 kHz.

Again, for freshwater habitats, no such general data are available and only a few studies have described the spectral profiles of ambient noise. Rivers commonly have high amounts of energy at low frequencies, sometimes followed by a "noise window" of relatively low noise levels in the 100–300 Hz region (Lugli and Fine, 2003) and a more or less steep rise of energy to about 1–2 kHz. The subsequent gradual decline with increasing frequency is also supported by our data. The Schwarza stream, however, differed by exhibiting a wider noise window (200–2000 Hz). Crawford *et al.* (1997) reported an acoustic window in the noise of the Niger River (Mali, Africa) in the 200–3000 Hz region, followed by high amounts of energies above 4 kHz, most likely of biotic origin (aquatic insects stridulating underwater)—a feature that was not observed in the present study. Apparently, the noise spectra of fast-flowing waters such as the Danube River in Austria, in contrast to the slowly flowing Niger River, are dominated by the water current, including transport of sediment and cavitation. This yields a considerable amount of energy in the high frequency range. Differences between flowing waters could mainly be attributed to differences in current velocity and type of bottom substrate.

Bom (1969) and Nystuen (1986) described very similar noise curves for lakes (Lake Sarzana, Italy, and Lake Clinton, IL., respectively), with high amounts of energy in low frequencies, followed by a rapid decline to higher frequencies and a more gradual decline above 10 kHz. We also observed the highest levels in the low-frequency region, followed by a fast decline, but only up to 1 kHz. Thereafter, the noise spectra remained relatively stable with only a gradual decline toward higher frequencies (Fig. 2). Bom (1969) and Nystuen (1986) also reported that these spectra depended highly on the weather conditions, especially rain. As we tried to measure under considerably good weather conditions, this can only be speculated for our data, but we expect the noise levels in Lake Neusiedl to be rather susceptible for wind-related noise due to its shallow water depth and large surface area.

The 6 to 12 kHz region of higher energy observed in two of the backwaters was due to aquatic insects. Aiken (1982) described spontaneous calls of a waterboatman (*Palmacorixa nana*, Corixidae), with dominant frequencies ranging from 5 to 8 kHz and Crawford *et al.* (1997) described for the Niger River that insect sounds considerably added to the noise floor above 4 kHz.

C. Effects of ambient noise on sound detection in fishes

While most studies on fish hearing have been performed under relatively quiet lab conditions, field studies in the sea (Loch Torridon, Scotland) showed that the hearing thresholds of four gadiform species (haddock *Melanogrammus aeglefinus*, pollack *Pollachius pollachius*, ling *Molva molva*, and cod *Gadus morhua*) were masked by the ambient noise (Champman, 1973; Champman and Hawkins, 1973). Several lab experiments (i.e., Fay, 1974; Amoser and Ladich, 2005; Wysocki and Ladich, 2005; Scholz and Ladich, 2006) have shown that the masking effect depends not only on level and type of noise but also on the overall hearing sensitivity of a given fish species.

In order to estimate the effect of the ambient noise encountered in the different habitats of this study, we correlated their amplitude spectra with those of habitat noise types analyzed by Amoser and Ladich (2005), whose effects on hearing are already known (Table III) with the assumption that highly correlated noise spectra in combination with comparable noise levels should have similar effects on hearing thresholds. Accordingly, we conclude that hearing thresholds measured under quiet lab conditions will only be reached in the quietest habitats such as Lake Lunz in fishes with enhanced hearing sensitivities (usually referred to as "hearing specialists," i.e., carps, minnows, catfishes) and that they will be moderately masked (by not more than 12 dB in their best hearing range) in the other stagnant habitats (lakes, backwaters, and pond). In contrast, they would be masked by up to 50 dB in the free-flowing areas of the Danube River, by 32 dB in creeks such as the Triesting, and by roughly 20 dB in slower-flowing parts of the Danube River such as in the Danube Channel and at the Danube Island. It is difficult to estimate auditory sensitivities in other creeks such as the Schwarza because its sound composition differed considerably from the Triesting stream, but the prevailing noise levels indicate that cyprinids are probably also significantly masked in this habitat. Perciforms (many prefer flowing habitats) would barely be masked in creeks (by maximally 9 dB) and even in the free-flowing Danube River at Orth (by only 13 dB).

Even though we recorded from only one site per habitat and there is certainly a variation within each habitat with regard to absolute levels and spectral composition in dependence on recording site and depth, season, or weather conditions, it is possible to draw the following general conclusions: (1) Absolute SPLs are clearly different between stagnant and fast-flowing habitats and (2) sound spectra from randomly selected sites per habitat are clearly correlated between similar habitat types while barely so between habitats with different limnological characteristics. Based on the observed differences in ambient noise level and composition, we conclude that improved absolute hearing sensitivity—as has been evolved in all cyprinids and catfishes by the development of an accessory peripheral auditory structure (the Weberian apparatus)—only represents a significant advantage compared to hearing nonspecialists (e.g., most perciforms) in quiet, standing waters. This advantage is limited in fast-flowing waters due to the high masking effects of the ambient noise.

One solution for optimizing intraspecific communication in a loud environment is to use acoustic "niches" which has been analyzed in three freshwater species living in quite different habitats. The African mormyrid *Pollimyrus isidori* in the Niger River as well as the European gobies *Padogobius martensii* and *Gobius nigricans* in the Italian Rivers Serchio and Stirone emit sounds with peak frequencies falling in the low frequency band of low ambient noise levels ("noise window") (Crawford *et al.*, 1997; Lugli and Fine, 2003). Our study confirmed that variable noise windows—regions of lower spectral energy in the low frequency range—seem to be a common feature in fast-flowing habitats, providing an acoustic niche for interception or communication. The extent to which such niches are utilized for intraspecific communication in the habitats analyzed in the present study is unknown. Three species from unrelated taxa are known to vocalize in these habitats: The marmoreal goby *Proterorhinus marmoratus* (Gobiidae; Ladich and Kratochvil, 1989), the sculpin *Cottus gobio* (Cottidae, Ladich, 1989), and the gudgeon *Gobio gobio* (Cyprinidae; Ladich, 1988). While the goby inhabits stagnant waters such as Danube River backwaters, the cyprinid and especially the sculpin prefer flowing waters such as the Triesting stream. The hearing specialist *G. gobio* will be masked in running waters, which is unlikely the case in the hearing generalist *C. gobio*.

In summary, noise levels as well as spectral composition are highly diverse in the various freshwater habitat types. Water currents, and the accompanying sediment transportation, are apparently the main cause of elevated noise levels in fast-flowing water systems; these specific hydrological and geological factors are probably also responsible for the relatively high variation of the spectral components of the habitat noise. In contrast, the sound spectra of stagnant water bodies are much more similar to each other, simplifying es-

timates of potential auditory sensitivity of autochthonous fish based on comparisons with previous studies. These data show that the environment significantly constrains the efficiency of a sensory system such as the auditory system, with enhanced auditory abilities presenting an advantage only in stagnant habitats with low noise levels. This provides additional evidence for the evolution of enhanced hearing capabilities in quiet freshwater habitats, as has been hypothesized in earlier studies (Schellart and Popper, 1992; Amoser and Ladich, 2005; Ladich and Popper, 2004).

ACKNOWLEDGMENTS

Aiken, R. B. (**1982**). "Sound production and mating in a waterboatman, *Palmacorixa nana* (Heteroptera: Corixidae)," Anim. Behav. **30**, 54–61.

Amoser, S., and Ladich, F. (**2005**). "Are hearing sensitivities of freshwater fish adapted to the ambient noise in their habitats?," J. Exp. Biol. **208**, 3533–3542.

Amoser, S., Wysocki, L. E., and Ladich, F. (**2004**). "Noise emission during the first powerboat race in an Alpine lake and potential impact on fish communities," J. Acoust. Soc. Am. **116**, 3789–3797.

Bom, N. (**1969**). "Effect of rain on underwater noise level," J. Acoust. Soc. Am. **45**, 150–156.

Boussard, A. (**1981**). "The reactions of roach (*Rutilus rutilus*) and rudd (*Scardinius erythrophthalmus*) to noises produced by high speed boating," *Proceedings of the second British Freshwater Fisheries Conference*, pp. 188–200.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer, Sunderland, MA).

Cato, D. H. (**1976**). "Ambient sea noise in waters near Australia," J. Acoust. Soc. Am. **60**, 320–328.

Chapman, C. J. (**1973**). "Field studies of hearing in teleost fishes," Helgoländer wissenschaftliche Meeresuntersuchungen **24**, 371–390.

Chapman, C. J., and Hawkins, A. D. (**1973**). "A field study of hearing in the cod, *Gadus morhua* L.," J. Comp. Physiol. **85**, 147–167.

Crawford, J. D., Jacob, P., and Bénech, V. (**1997**). "Sound production and reproductive ecology of strongly acoustic fish in Africa: *Pollimyrus isidori*, Mormyridae," Behaviour **134**, 677–725.

Egner, S. A., and Mann, D. A. (**2005**). "Auditory sensitivity of sergeant major damselfish *Abudefduf saxatilis* from post-settlement juvenile to adult," Mar. Ecol.: Prog. Ser. **285**, 213–222.

Fay, R. R. (**1974**). "Masking of tones by noise for the goldfish," J. Comp. Physiol. Psychol. **87**, 708–716.

Fine, M. L., and Lenhardt, M. L. (**1983**). "Shallow water propagation of the toadfish mating call," Comp. Biochem. Physiol. A **76**, 225–231.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Gerald, J. W. (**1971**). "Sound production in six species of sunfish (Centrarchidae)," Evolution (Lawrence, Kans.) **25**, 75–87.

Hawkins, A. D., and Johnstone, A. D. F. (**1978**). "The hearing of the Atlantic Salmon, *Salmon salar*," J. Fish Biol. **13**, 655–673.

Hawkins, A. D., and Myrberg, A. A., Jr. (**1983**). "Hearing and sound communication under water," in *Bioacoustics, a Comparative Approach*, edited by B. Lewis (Academic, London), pp. 347–405.

ISO 1996 (**2003**). "Acoustics: Description, measurement and assessment of environmental noise," International Organization for Standardization

Ladich, F. (**1988**). "Sound production by the gudgeon, *Gobio gobio* L.; a common European freshwater fish (Cyprinidae, Teleostei)," J. Fish Biol. **32**, 707–715.

Ladich, F. (**1989**). "Sound production by the river bullhead *Cottus gobio* L. (Cottidae, Teleostei)," J. Fish Biol. **35**, 531–538.

Ladich, F., and Bass, A. H. (**2003**). "Underwater sound generation and acoustic reception in fishes with some notes on frogs," in *Sensory Processing in Aquatic Environments*, edited by S. P. Collin and N. J. Marshall (Springer, New York), pp. 173–193.

Ladich, F., and Kratochvil, H. (**1989**). "Sound production by the marmoreal goby, *Protherorhinus marmoratus* (Pallas) (Gobiidae, Teleostei)," Zoologische Jahrbücher. Abteilung für allgemeine Zoologie und Physiologie der Tiere **93**, 501–504.

Ladich, F., and Popper, A. N. (**2004**). "Parallel evolution in fish hearing organs," in *Evolution of the Vertebrate Auditory System*, edited by G. A. Manley, A. N. Popper, and R. R. Fay (Springer, New York), pp. 95–127.

Lugli, M., and Fine, M. L. (**2003**). "Acoustic communication in two freshwater gobies: Ambient noise and short-range propagation in shallow streams," J. Acoust. Soc. Am. **114**, 512–521.

Lugli, M., Yan, H. Y., and Fine, M. L. (**2003**). "Acoustic communication in two freshwater gobies: The relationship between ambient noise, hearing thresholds and sound spectrum," J. Comp. Physiol., A **189**, 309–320.

Mann, D. A. (**2006**). "Propagation of fish sounds," in *Communication in Fishes*, edited by F. Ladich, S. P. Collin, P. Moller, and B. G. Kapoor, Vol. **1** (Science Publishers, Enfield, NH), pp. 107–120.

Mann, D. A., and Lobel, P. S. (**1997**). "Propagation of damselfish (Pomacentridae) courtship sounds," J. Acoust. Soc. Am. **101**, 3783–3791.

McConnell, S. O., Schilt, M. P., and Dworski, J. G. (**1992**). "Ambient noise measurements from 100 Hz to 80 kHz in an Alaskan fjord," J. Acoust. Soc. Am. **91**, 1990–2003.

McDonald, M. A., Hildebrand, J. A., and Wiggins, S. M. (**2006**). "Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California," J. Acoust. Soc. Am. **120**, 171–718.

Nystuen, J. A. (**1986**). "Rainfall measurements using underwater ambient noise," J. Acoust. Soc. Am. **79**, 972–981.

Richardson, W. J., Greene, C. R., Malme, C. J., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Samuel, Y., Morreale, S. J., Clark, C. W., Greene, C. H., and Richmond, M. E. (**2005**). "Underwater, low-frequency noise in a coastal sea turtle habitat," J. Acoust. Soc. Am. **117**, 1465–1472.

Schellart, N. A. M., and Popper, A. N. (**1992**). "Functional aspects of the evolution of the auditory system of actinopterygian fish," in *The Evolutionary Biology of Hearing*, edited by D. E. Webster, R. R. Fay, and A. N. Popper (Springer, New York), pp. 295–322.

Scholz, K., and Ladich, F. (**2006**). "Sound production, hearing and possible interception under ambient noise conditions in the topmouth minnow *Pseudorasbora parva*," J. Fish Biol. **69**, 692–906.

Tavolga, W. N. (**1967**). "Masked auditory thresholds in teleost fishes," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, Oxford, UK), pp. 233–245.

Tolimieri, N., Haine, O., Jeffs, A., McCauley, R, D., and Montgomery, J, C. (**2004**). "Directional orientation of pomacentrid larvae to ambient reef sound," Coral Reefs **23**, 184–191.

Townend, J. (**2002**). *Practical statistics for environmental and biological scientists* (Wiley, Chichester, UK), pp. 132–134.

Urick, R, J. (**1983**). "The noise background of the sea: ambient noise level," in *Principles of Underwater Sound*, edited by R. J. Urick (Peninsula, Los Altos, CA), pp. 202–236.

Wenz, G. M. (**1962**). "Acoustic ambient noise in the ocean: spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956.

Wood, J. D., McCowan, B., Langbauer, W. R., Jr., Viljoen, J. J., and Hart, L. A. (**2005**). "Classification of African elephant *Loxodonta africana* rumbles using acoustic parameters and cluster analysis," Bioacoustics 15, 143–161.

Wysocki, L. E., and Ladich, F. (**2005**). "Hearing in fishes under noise conditions," J. Assoc. Res. Otolaryngol. **6**, 28–36.

Yamazaki, Y., Haramoto, S., and Fukasawa, T. (**2006**). "Habitat uses of freshwater fishes on the scale of reach system provided in small streams," Environmental Biology of Fishes **75**, 333–341.

Bayesian focalization: Quantifying source localization with environmental uncertainty

Stan E. Dosso[a] and Michael J. Wilmut

School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

This paper applies a Bayesian formulation to study ocean acoustic source localization as a function of uncertainty in environmental properties (water column and seabed) and of data information content [signal-to-noise ratio (SNR) and number of frequencies]. The approach follows that of the optimum uncertain field processor [A. M. Richardson and L. W. Nolte, J. Acoust. Soc. Am. **89**, 2280–2284 (1991)], in that localization uncertainty is quantified by joint marginal probability distributions for source range and depth integrated over uncertain environmental properties. The integration is carried out here using Metropolis Gibbs' sampling for environmental parameters and heat-bath Gibbs' sampling for source location to provide efficient sampling over complicated parameter spaces. The approach is applied to acoustic data from a shallow-water site in the Mediterranean Sea where previous geoacoustic studies have been carried out. It is found that reliable localization requires a sufficient combination of prior (environmental) information and data information. For example, sources can be localized reliably for single-frequency data at low SNR (−3 dB) only with small environmental uncertainties, whereas successful localization with large environmental uncertainties requires higher SNR and/or multifrequency data. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2715451]

I. INTRODUCTION

Matched-field source localization in shallow-water environments is an important problem that has received considerable attention in recent years.[1–3] Matched-field methods estimate source range and depth by matching (correlating) acoustic fields measured at an array of sensors with replica fields computed for a grid of possible source locations using a numerical propagation model. The results are typically displayed as an ambiguity surface, consisting of the computed matches plotted as a function of range and depth. The source location is estimated as the position of maximum match, with the structure of local maxima (sidelobes of the correlator) providing a qualitative indication of the nonuniqueness/uncertainty of this estimate. Localization requires good knowledge of the physical properties of the ocean environment (water column and seabed),[4] which strongly affect the propagation of acoustic fields. However, environmental properties may not be well known, and, in practice, environmental uncertainty often represents the limiting factor in matched-field localization. Several approaches have been developed to address this limitation, including geoacoustic inversion,[5–18] focalization,[19] and the optimal uncertain field processor.[20–23]

Geoacoustic inversion surveys can be carried out for particular regions with the goal of inverting acoustic fields measured for a controlled source to determine seabed parameters appropriate for subsequent localization of unknown sources. Although the source range and depth are approximately known in the survey, they are usually included in the inversion as unknown parameters with small search bounds to account for experimental uncertainties. Geoacoustic inversion is a strongly nonlinear problem, and considerable effort has been applied to develop effective approaches to determine optimal solutions using global search methods.[5–10] More recently, nonlinear Bayesian methods have been applied to estimate parameter uncertainties in geoacoustic inversion using Markov-chain Monte Carlo sampling.[11–14] Although the recovered geoacoustic properties are usually treated as known (fixed) parameters in subsequent source localizations, their inherent uncertainties ultimately translate to localization uncertainties. Further, although water-column sound-speed profiles (SSP) are more straightforward to measure than seabed parameters, uncertainty in SSP due to temporal and/or spatial variability can also significantly affect source localization.[15,16]

The method of focalization[19] addresses environmental uncertainty in matched-field localization by including the environmental parameters as additional unknowns in an augmented localization problem which is solved using a global search. In principle, focalization and geoacoustic inversion address the same problem (joint inversion for source location and environmental parameters), but with different prior information: in geoacoustic inversion the source location is relatively well known *a priori*, while in focalization it is not. This difference is significant, as increasing the search bounds for source location beyond a small correction typically admits a large number of local optima to the parameter space. Given this increased complexity, the goal of focalization is to obtain effective environmental parameters which provide improved localization, but may not represent the actual environment.[19] Standard focalization does not provide a

[a]Electronic mail: sdosso@uvic.ca

measure of localization uncertainty (e.g., no ambiguity surface).

The optimum uncertain field processor (OUFP)[20–23] represents a Bayesian approach to source localization with environmental uncertainty. In a Bayesian formulation, the solution to an inverse problem is characterized by its posterior probability density (PPD), which combines prior information of the unknown parameters with information from observed data expressed in terms of a likelihood function. OUFP is based on integrating the multidimensional PPD over uncertain environmental parameters to produce a joint marginal probability distribution over source range and depth, referred to as a probability ambiguity surface (PAS). The maximum *a posteriori* (MAP) estimate from this distribution represents the most probable source location, and the distribution itself provides a quantitative measure of localization uncertainty. In OUFP, the integration is carried out by Monte Carlo sampling drawn from the prior distribution for environmental parameters.

This paper applies the concept of Bayesian localization with environmental uncertainty developed in OUFP. However, a different formulation is developed, representing a generalization of Bayesian geoacoustic inversion approaches.[12–14] In particular, the integration over uncertain parameters is carried out here using Markov-chain Monte Carlo importance sampling methods which sample directly from the PPD, rather than from the prior distribution. Specifically, the method of fast Gibbs sampling,[12,13] which applies adaptive Metropolis Gibbs sampling in a principal-component parameter space, is applied to geoacoustic and SSP parameters, whereas heatbath Gibbs sampling[24] is applied to source range and depth to provide efficient sampling over complicated source search spaces which can include many isolated local optima. A rigorous convergence criterion is applied for the numerical integration to ensure reliable results. Although OUFP integrates over unknown source amplitude and phase (assumed Rayleigh and uniform distributed, respectively), the approach here applies analytic maximum-likelihood estimates for amplitude and phase. Because our approach integrates over uncertain environmental parameters in an augmented localization problem, but with a different formulation than OUFP, the method is referred to here as Bayesian focalization. The Bayesian focalization approach treats data uncertainties as unknown within the inversion using an efficient implicit scheme (see Sec. II). Finally, the efficient Gibbs sampling of Bayesian focalization also provides environmental uncertainties in the form of marginal distributions, correlations, etc., for complex environments (12 unknown parameters representing the water column and seabed in the examples presented here).

In this paper, Bayesian focalization is applied to acoustic data recorded at a vertical hydrophone array at a shallow-water test site in the Mediterranean Sea where previous geoacoustic inversion surveys have been carried out.[14,25] The goals of the present analysis are to examine localization uncertainties as a function of prior information (i.e., *a priori* uncertainties in geoacoustic and SSP parameters) and of data information content [signal-to-noise ratio (SNR) and number of frequencies included in the inversion].

II. THEORY

This section summarizes the Bayesian focalization formulation; more complete treatments of Bayesian inference can be found elsewhere.[24,26,27] Let $\mathbf{m}$ and $\mathbf{d}$ represent vectors of model parameters (environmental properties and source location) and data (measured acoustic fields), respectively, with elements considered to be random variables. Bayes' rule may be expressed

$$P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{m})P(\mathbf{m}), \qquad (1)$$

where $P(\mathbf{m}|\mathbf{d})$ represents the PPD and $P(\mathbf{m})$ is prior information. Interpreting $P(\mathbf{d}|\mathbf{m})$ as a function of $\mathbf{m}$ for the measured data defines the likelihood function, which can generally be expressed $L(\mathbf{m}) \propto \exp[-E(\mathbf{m})]$, where $E(\mathbf{m})$ is the data misfit function (considered below). The PPD can then be written

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-\phi(\mathbf{m})]}{\int \exp[-\phi(\mathbf{m}')]d\mathbf{m}'}, \qquad (2)$$

where

$$\phi(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m}) \qquad (3)$$

is a generalized misfit function including both data and prior, and the domain of integration spans the parameter space. To interpret the multidimensional PPD requires computation of integral properties, such as the posterior mean, model covariance matrix, and marginal probability distributions, defined, respectively, as

$$\bar{\mathbf{m}} = \int \mathbf{m} P(\mathbf{m}|\mathbf{d})d\mathbf{m}, \qquad (4)$$

$$\mathbf{C} = \int (\mathbf{m} - \bar{\mathbf{m}})(\mathbf{m} - \bar{\mathbf{m}})^T P(\mathbf{m}|\mathbf{d})d\mathbf{m}, \qquad (5)$$

$$P(m_i|\mathbf{d}) = \int \delta(m_i' - m_i)P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \qquad (6)$$

where δ is the Dirac delta function and joint marginal distributions are defined in a manner similar to Eq. (6). Parameter uncertainties can also be quantified in terms of credibility intervals, such as the $\beta\%$ highest-probability density (HPD) interval, representing the interval of minimum width containing $\beta\%$ of the area of a marginal distribution. Interparameter correlations are quantified by normalizing the covariance matrix to produce the correlation matrix

$$S_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}. \qquad (7)$$

Elements S_{ij} are within $[-1, +1]$, with a value of $+1$ (-1) indicating perfect correlation (anticorrelation) between m_i and m_j.

For nonlinear problems, such as matched-field inversion, the previous integrals must be solved numerically, and Markov-chain Monte Carlo methods such as Gibbs' sampling (GS) are typically employed.[18–20] In this paper, two general forms of GS are applied. In Metropolis GS, each parameter of the model is perturbed randomly in turn, with

perturbations accepted according to the Metropolis criterion; i.e., if a uniform random number ξ drawn from the interval $[0,1]$ satisfies

$$\xi \leq \exp[-\Delta\phi(\mathbf{m})]. \tag{8}$$

The efficiency of Metropolis GS can be improved by applying perturbations to rotated parameters in a principal-component parameter space obtained by diagonalizing an estimate of the model covariance matrix, Eq. (5), from early sampling.[13,14] In heat-bath GS, a discrete (unnormalized) probability distribution is formed for each parameter m_i in turn by sampling $\exp[-\phi(\mathbf{m})]$ across the parameter's range while holding other parameters fixed. A new value for m_i is then drawn at random from this distribution and accepted unconditionally. Markov-chain analysis verifies that, in the limit of a large number of perturbations, both Metropolis and heat-bath GS provide an unbiased sampling of the PPD, Eq. (2).[24,27]

Although Metropolis GS has proved effective for geoacoustic inversion,[12–14] this approach was found to be inefficient for sampling source-location parameters. The difficulty results from the typically large number of isolated, locally optimal solutions in source range and depth (r and z) without dominant correlation patterns. As a result, one-dimensional sampling in r and in z is inefficient, since the probability of jumping between isolated, nonaligned optima in two dimensions is relatively small. However, two-dimensional heat-bath sampling in the r–z plane precludes this difficulty. In this approach, joint marginal probability distributions for source location are formed by summing the r–z distributions constructed for a large number of realizations of the environmental parameters as generated using Metropolis GS with parameter rotation (i.e., averaging r–z probability surfaces over environmental parameters sampled from the PPD). For range-independent problems, r–z probability surfaces can be computed very efficiently using normal-mode models, as the modal properties are defined by the environment not the source location (i.e., modal properties need be computed only once per environmental realization).[28]

Convergence of the inversion algorithm is judged in terms of the difference between two independent samples of models collected in parallel. In particular, convergence requires the maximum difference between the two independent two-dimensional cumulative marginal distributions in r and z be sufficiently small (e.g., <0.02). The final sample for integration is then taken to be the union of the two independent samples.

To define the data misfit function, $E(\mathbf{m})$, consider complex acoustic-field data measured at an array of N sensors at F frequencies, $\mathbf{d} = \{\mathbf{d}_f, f=1, F\}$. Assuming the data errors are complex, zero mean, Gaussian-distributed random variables uncorrelated over space and frequency with variance ν_f at the fth frequency, the likelihood function is given by

$$L(\mathbf{m}) = \prod_{f=1}^{F} \frac{1}{(\pi\nu_f)^N} \exp[-|\mathbf{d}_f - A_f e^{i\theta_f}\mathbf{d}_f(\mathbf{m})|^2/\nu_f], \tag{9}$$

where $\mathbf{d}_f(\mathbf{m})$ is the replica acoustic field computed for model $\mathbf{m}$ via a numerical propagation model and $A_f e^{i\theta_f}$ represents
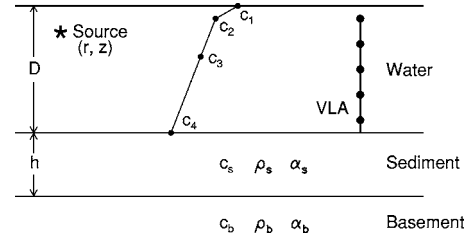
FIG. 1. Schematic diagram of experiment and parameterization including geoacoustic, SSP, and geometric parameters.

the source magnitude and phase. In localization applications for unknown sources, the source spectrum and data variances are generally not known, and can be treated implicitly by maximizing the likelihood function over A_f, θ_f, and ν_f (i.e., setting $\partial L/\partial A_f = \partial L/\partial \theta_f = \partial L/\partial \nu_f = 0$), yielding the misfit function[17]

$$E(\mathbf{m}) = N \sum_{f=1}^{F} \log_e B_f(\mathbf{m}), \tag{10}$$

where $B_f(\mathbf{m})$ represents the (unnormalized) Bartlett mismatch defined

$$B_f(\mathbf{m}) = |\mathbf{d}_f|^2 - \frac{|\mathbf{d}_f(\mathbf{m})^\dagger \mathbf{d}_f|^2}{|\mathbf{d}_f(\mathbf{m})|^2}. \tag{11}$$

III. EXPERIMENT, DATA, AND MODEL

The PROSIM'97 shallow-water geoacoustic experiment was carried out by the SACLANT (now NATO) Undersea Research Centre in the Mediterranean Sea off the west coast of Italy near Elba Island. The experiment and data have been considered previously,[14,25] and will be described only briefly here. The experiment consisted of recording acoustic signals from a transducer towed at approximately 10 m depth over a series of shallow-water tracks. The source emitted a 0.5 s linear frequency-modulated signal over the band 300–800 Hz every ~0.25 km along the track. The signals were received at a bottom-moored vertical line array (VLA) of 48 hydrophones which spanned from 26 to 120 m depth with a 2 m sensor spacing. The data analyzed here were recorded along a track with nearly range-independent bathymetry (water depth ~132 m). SSPs measured during the experiment were found to be stable both spatially and temporally, consisting of a weakly downward-refracting gradient that varied from about 1520 to 1510 m/s over the water column.

The environment and source parameters that comprise the model $\mathbf{m}$ for Bayesian focalization are illustrated in Fig. 1. The acoustic source is at depth z and range r from the VLA in water of depth D. The SSP is represented by four sound-speed parameters c_1–c_4 at depths of 0, 10, 50, and D m. The geoacoustic parameters include the thickness h of an upper sediment layer with sound speed c_s, density ρ_s, and attenuation α_s, overlying a semi-infinite basement with sound speed c_b, density ρ_b, and attenuation α_b.

Geoacoustic inversion was applied previously to the PROSIM'97 data for a source range of approximately 3.95 km, employing 11 frequencies at 50 Hz intervals over

TABLE I. Summary of environmental parameters and bounds. For geoacoustic parameters, best estimates represent the optimal solution from matched-field inversion in Ref. 14, with wide bounds representing prior uncertainties and narrow bounds representing 95% HPD credibility intervals (applied to Bayesian focalization in this paper). For SSP parameters, best estimates represent measured values with arbitrary wide and narrow bounds applied in Bayesian focalization.

Parameter and Units	Best estimate	Wide bounds	Narrow bounds
Geoacoustics			
h (m)	9.0	[0, 30]	[7, 11]
c_s (m/s)	1494	[1450, 1600]	[1486, 1504]
c_b (m/s)	1529	[1500, 1650]	[1524, 1531]
ρ_s (g/cm^3)	1.38	[1.0, 1.7]	[1.3, 1.7]
ρ_b (g/cm^3)	1.52	[1.5, 2.2]	[1.5, 2.1]
α_s (dB/λ)	0.02	[0, 1]	[0.0, 0.2]
α_b (dB/λ)	0.22	[0, 1]	[0.0, 0.5]
D (m)	130.1	[128, 135]	[130.0, 132.5]
SSP			
c_1 (m/s) at 0 m	1520	[1515, 1525]	[1519, 1521]
c_2 (m/s) at 10 m	1517	[1510, 1520]	[1516, 1528]
c_3 (m/s) at 50 m	1515	[1510, 1520]	[1514, 1516]
c_4 (m/s) at 130 m	1510	[1505, 1515]	[1509, 1511]

FIG. 3. (Color online) PASs for various source ranges (indicated by column labels) and levels of prior information (row labels) including combinations of narrow and wide bounds for geoacoustic and SSP parameters as defined in Table I. Dotted lines indicate the true source depth and range. Acoustic data are single frequency (300 Hz) with SNR=−3 dB.

IV. RESULTS

A. Effect of prior information

This section considers the dependence of source localization uncertainty distributions on the prior information (uncertainties) of environmental parameters. To this end, Fig. 3 compares PASs computed for measured data at source ranges of approximately 3.1, 4.2, 5.3, and 6.3 km, with four different states of environmental information consisting of uniform prior distributions with: wide prior bounds for both geoacoustic and SSP parameters (top row of Fig. 3); wide bounds for geoacoustic parameters and narrow bounds for SSP parameters (second row); narrow bounds for geoacoustic parameters and wide bounds for SSP parameters (third row); and narrow bounds for both geoacoustic and SSP parameters (bottom row). The wide and narrow bounds for all parameters are given in Table I. The wide geoacoustic bounds are identical with the prior bounds used in the earlier geoacoustic inversion,[14] and represent poor knowledge of the seabed properties. The narrow geoacoustic bounds are taken to be the 95% HPD credibility interval from the geoacoustic inversion results shown in Fig. 2, and represent the state of seabed information typically available for localization from a previous geoacoustic inversion survey in a given region. The wide and narrow SSP bounds consist of intervals 10 and 2 m/s wide, respectively, with the measured sound-speed values at or near the center of the interval. To avoid applying localization to data sets that are too similar to those used for the geoacoustic inversion and to more realistically represent typical localization scenarios, the acoustic data used in the Bayesian focalization at each range consist of the complex pressure recorded at every second hydrophone of the array (i.e., 24 sensors at 4 m intervals) at a single frequency of 300 Hz, with random Gaussian-distributed errors added to the measured data to reduce the SNR to −3 dB to more realistically represent an unknown, low-level source (identical errors were applied for each source range in Fig. 3). The definition of SNR used here is

the 300–800-Hz source band.[14] The SNR for the data was approximately 30 dB, although the effective SNR,[14] which also accounts for theory error, varied from about 7 to 0 dB over the band. The seabed and geometric parameters shown in Fig. 1 were included in the geoacoustic inversion, while the SSP parameters were held fixed at their measured values (given in Table I). The prior information for the geoacoustic parameters consisted of uniform distributions over wide parameter bounds representing poor knowledge of seabed properties for this environment (Table I). The geoacoustic inversion results[14] are given as marginal probability distributions in Fig. 2, and as MAP parameter estimates and 95% HPD credibility intervals in Table I.

FIG. 2. Marginal probability distributions for seabed parameters from geoacoustic inversion in Ref. 14.

S. E. Dosso and M. J. Wilmut: Bayesian focalization

FIG. 5. (Color online) (a) PAS for source at 4.2 km range from Bayesian focalization over narrow prior uncertainty bounds on geoacoustic and SSP parameters. (b)–(d) PASs for fixed environmental parameters drawn at random from narrow uncertainty bounds (no integration).

FIG. 4. Depth and range error of PAS peak and IP computed from PASs (Fig. 3) for various levels of prior information including combinations of narrow and wide bounds for geoacoustic and SSP parameters as indicated. Acoustic data are single frequency (300 Hz) with SNR$=-3$ dB. Results for source ranges of 3.1, 4.2, 5.3, and 6.3 km are represented by thin solid, dashed, dotted, and thick solid lines, respectively.

$$\text{SNR}_f = 10 \log_{10} \frac{|\mathbf{p}_f|^2}{|\mathbf{n}_f|^2}, \tag{12}$$

where $\mathbf{p}_f$ and $\mathbf{n}_f$ represents the measured signal and additive noise vector at the fth frequency. Note that, in addition to the additive Gaussian errors, the data uncertainties also include theory error resulting from the limitations of the model parameterization and forward modeling (the normal-mode model ORCA[29] was used to generate replica fields for focalization).

For display purposes, the PAS plots in Fig. 3 can be normalized in several ways to emphasize different features in their comparison; the choice in Fig. 3 is that a common normalization is applied for each source range (i.e., each column of Fig. 3 has a common normalization). Some features of the PASs in Fig. 3 are quantified further in Fig. 4, which shows the the depth error (absolute difference between true source depth and PAS peak), range error (defined similarly), and the integrated probability (IP) for an acceptable localization, defined to be the integral of the PAS over an acceptable region taken to be within ±10 m depth and ±300 m range of the true source position.

Figures 3 and 4 show that a region of at least slightly elevated probability (relative to the background level) is associated with the true source location in all cases. For instance, if the probability was uniformly distributed over the 8 km×120 m $r-z$ source search region, the IP of the 0.6 km×20 m acceptable region would be 0.0125, whereas Fig. 4 shows that the IP values are ≳0.1 for all cases. However, localization results are generally poor for the two cases involving wide geoacoustic parameter bounds, with significant range errors and low IP values, although good depth estimates are obtained for three source ranges with wide geoacoustic and narrow SSP bounds. Localization is significantly improved for the case of narrow geoacoustic and wide SSP bounds, with small depth and range errors for three of
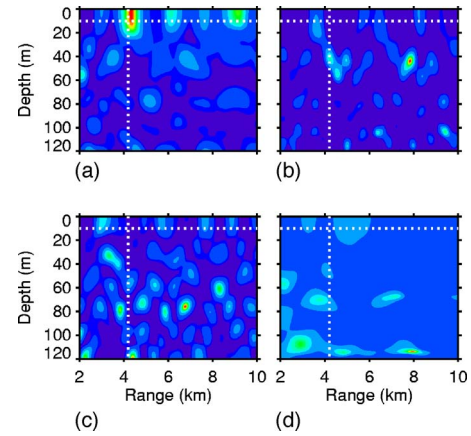
four source ranges and IP values of approximately 0.3. Finally, for narrow geoacoustic and SSP bounds, Bayesian focalization provides good localization for all four source ranges, with an average IP value of about 0.5.

It is interesting to note in Fig. 3 that wide environmental bounds can lead to a high-probability PAS peak at an incorrect source location (e.g., 4.2 km range, wide geoacoustic bounds). This spurious peak results from a fortuitous combination of parameters that produces a good fit to a particular noisy data set. This peak is substantially reduced when narrower prior bounds are applied. Further, the spurious peak was not observed when other random noise realizations were applied, whereas the PAS peak at the true source location for narrow environmental bounds was found to persist for other noise realizations (not shown).

It is important to note that the good localization results obtained for the narrow geoacoustic and SSP prior bounds in Figs. 3 and 4 do not indicate that any and all environmental models within these bounds will suffice for localization. Rather, the Bayesian focalization samples specific parameter combinations within the prior distribution that are consistent with the acoustic data (within uncertainties) in integrating over environmental parameters. To illustrate this point, Fig. 5 compares the PAS for the source at 4.2 km range computed via GS integration over narrow environmental bounds to PASs computed for three cases in which geoacoustic and SSP parameters were drawn at random from the narrow bounds and held fixed (i.e., no environmental integration is carried out). Each of the three fixed-environment cases yield very poor localization results.

B. Effect of data information

This section considers the dependence of source localization uncertainty distributions on data information content, in particular, on the data quality (in terms of SNR) and on the data quantity (number of frequencies included in multifrequency localization). In all cases considered in this section, wide prior bounds are applied for geoacoustic and SSP parameters, representing source localization in a poorly unknown environment.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

S. E. Dosso and M. J. Wilmut: Bayesian focalization   2571
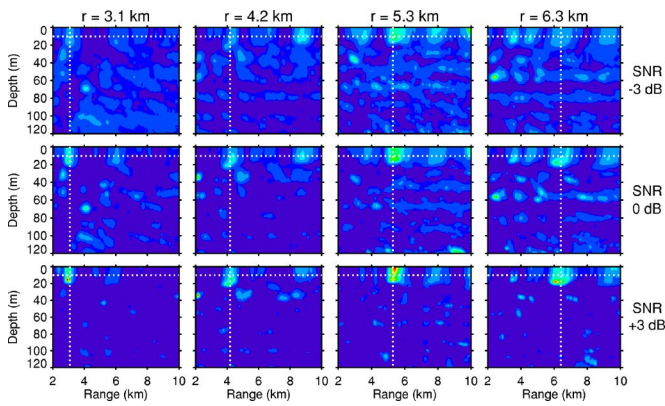
FIG. 6. (Color online) PASs for various source ranges (indicated by column labels) and SNRs (row labels). Dotted lines indicate the true source depth and range. Acoustic data are single frequency (300 Hz), and wide prior bounds are applied for geoacoustic and SSP parameters.

FIG. 8. (Color online) PASs for various source ranges (indicated by column labels) and frequencies included in inversion (row labels). Dotted lines indicate the true source depth and range. Wide prior bounds are applied for geoacoustic and SSP parameters, and SNR=−3 dB.

The effect of SNR on Bayesian focalization is examined in Figs. 6 and 7 by considering PASs for the same four source ranges considered previously with random Gaussian errors added to the measured 300 Hz acoustic data to achieve SNRs of −3, 0, and +3 dB. Figures 6 and 7 indicate an elevated probability level at the true source location in all cases. However, the global probability maxima do not coincide with the true source locations for any of the four source ranges at SNR=−3 dB. At 0 dB, the probability maxima agree with the true locations for two out of four source ranges, and at +3 dB agreement is achieved for three of four ranges. The integrated probability over the acceptable region increases from ∼0.1 at −3 dB to as much as 0.8 at +3 dB for the correctly-localized sources. Even for the 4.2 km source that is not correctly localized, the integrated probability about the true location increases to >0.4 at +3 dB, indicating that this is a highly significant local maximum. Overall, these results indicate that with sufficiently high data quality (i.e.,

high enough SNR), Bayesian focalization has the potential for high-probability localization even with large prior uncertainties in environmental parameters.

The effect of the number of acoustic frequencies on Bayesian focalization is examined in Figs. 8 and 9 by considering PASs for data with SNR_f=−3 dB at each frequency for cases with one, two, and three frequencies included in the inversion (300; 300 and 350; and 300, 350, and 400 Hz, respectively). As before, elevated probability levels are obtained at the true source location in all cases, although correct localization is not achieved for any source range with a single data frequency. However, correct localization is achieved for all four source ranges for the cases of two and three frequencies. For three of the four source ranges, the integrated probability of acceptable localization increases significantly in going from one to two frequencies and remains roughly constant from two to three frequencies. For

FIG. 7. Depth and range error of PAS peak and IP computed from PASs (Fig. 6) for various SNRs as indicated. Acoustic data are single frequency (300 Hz) and wide prior bounds are applied for geoacoustic and SSP parameters. Results for source ranges of 3.1, 4.2, 5.3, and 6.3 km are represented by thin solid, dashed, dotted, and thick solid lines, respectively.
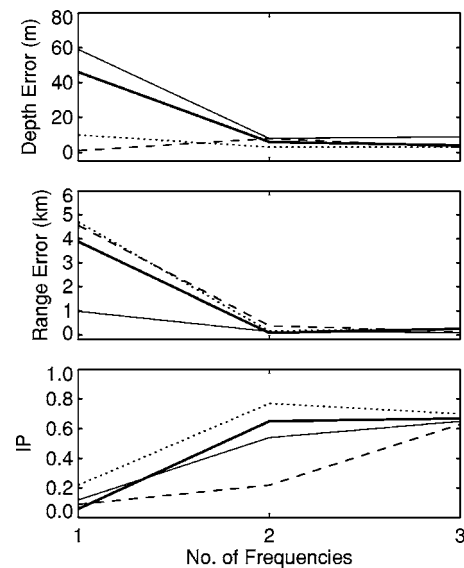
FIG. 9. Depth and range error of PAS peak and IP computed from PASs (Fig. 8) for various numbers of frequencies included in the inversion, as indicated. Wide prior bounds are applied for geoacoustic and SSP parameters, and SNR=−3 dB. Results for source ranges of 3.1, 4.2, 5.3, and 6.3 km are represented by thin solid, dashed, dotted, and thick solid lines, respectively.
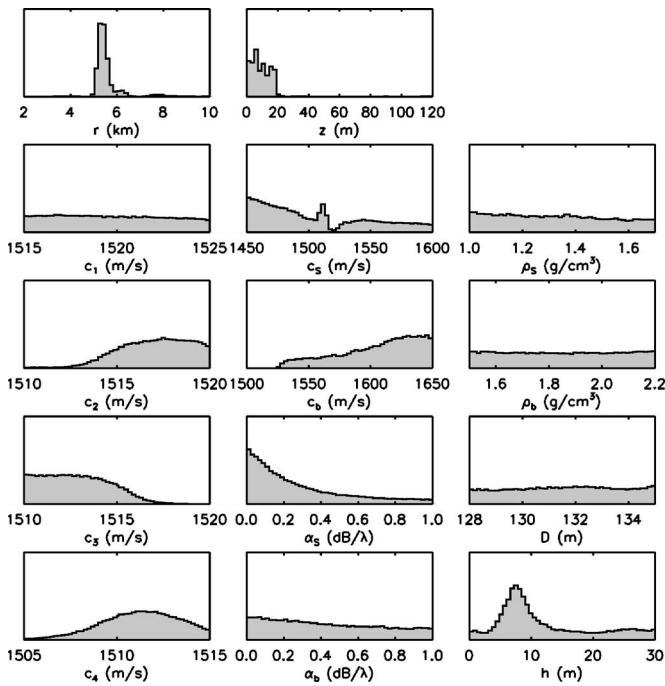
FIG. 10. Marginal probability distributions for source and environmental parameters from Bayesian focalization of 5.3 km source (300 Hz data with SNR=+3 dB). Horizontal plot limits indicate the prior bounds for all parameters. The vertical scale for source location parameters (r and z) is three times that for other parameters.

the 4.2 km source, the IP value shows a smaller increase in going from one to two frequencies, but increases significantly from two to three frequencies. In all cases, IP values of 0.6–0.7 are achieved for three-frequency focalization, indicating strong global maxima. These results quantify the substantial improvement in source localization possible with multiple-frequency acoustic data, with substantial benefits for even two frequencies.

Although the primary goal of this work is the construction of PASs to quantify source localization uncertainty, it is interesting to consider the extent to which the environmental parameters are resolved in the process of Bayesian focalization for an unknown source. Figure 10 shows marginal probability distributions for all parameters computed via Bayesian focalization for 300 Hz acoustic data for the source at 5.3 km range with SNR=+3 dB. The marginal PPDs for r and z show high probabilities associated with the correct source location. However, the marginal distributions for the environmental parameters indicate poor resolution and generally extend over the entire prior bounds, with a few notable exceptions. For instance, marginal distributions for c_2-c_4 indicate some ability to resolve these SSP parameters within their bounds, while the marginal PPD for c_1 is uniform over its bounds, likely because this parameter effects only a thin water layer (0–10 m depth). The marginal PPD for sediment thickness, h, shows a strong peak near 8 m, consistent with the geoacoustic inversion results in Fig. 2 although the distribution tails extend over the entire bounds. The marginal PPD for sediment attenuation, α_s, indicates a low value, consistent with Fig. 2. Finally, the basement sound speed, c_b, although poorly resolved, is constrained to be >1520 m/s, again consistent with Fig. 2.
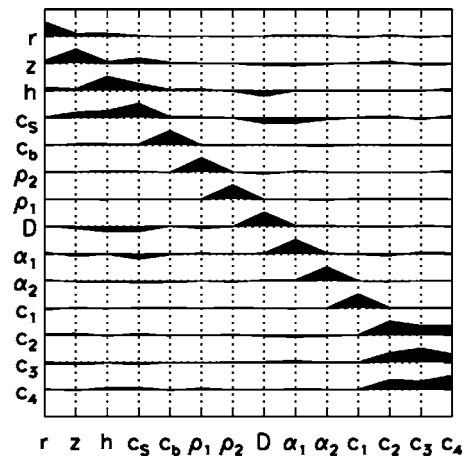
FIG. 11. Correlation matrix for source and environmental parameters from Bayesian focalization of 5.3 km source (300 Hz data with SNR=+3 dB).

To consider interparameter relationships, Fig. 11 shows the correlation matrix for this inversion. Correlations between geoacoustic parameters such as sediment thickness h and sound speed and c_s, and between h and water depth D have been discussed elsewhere.[13,14] Of note here are the strong positive correlations between SSP parameters c_2-c_4. These correlations indicate that the general shape of the SSP (below the surficial layer) is constrained in the focalization, even though the actual SSP values are not. To illustrate this point, Fig. 12 compares the measured SSP to 20 SSPs randomly chosen from the PPD sample. All of the SSP samples reflect the general downward-refracting character (decreasing sound speed with depth) of the measured SSP below 10 m depth, although the surficial sound-speed value is poorly constrained with both positive and negative gradients in the thin upper-most layer.

Finally, it is worth emphasizing that the large posterior uncertainties in environmental parameters obtained here do not indicate that just any environmental model within the bounds will suffice for localization. As in Fig. 5, arbitrarily chosen (fixed) environmental models were found to produce very poor localization results.
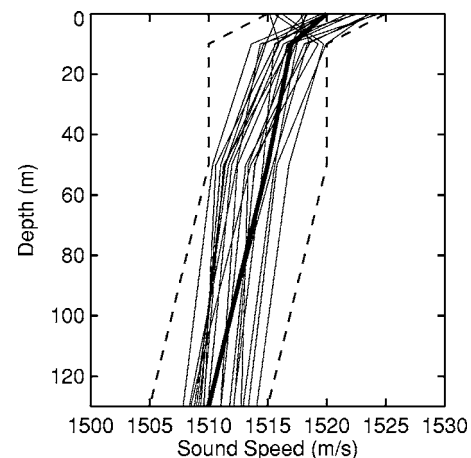
FIG. 12. Random selection of 20 SSPs from the PPD computed via Bayesian focalization of 5.3 km source (300 Hz data with SNR=+3 dB). Thick line represents the measured SSP and dashed lines indicate prior SSP parameter bounds.

V. SUMMARY

This paper applied a Bayesian approach to the problem of matched-field source localization with environmental uncertainty. The approach follows that of the optimum uncertain field processor,[20–23] in that localization uncertainty is quantified by the joint marginal probability distribution in source range and depth (referred to as a probability ambiguity surface), with the dependence on unknown environmental parameters integrated out. The formulation developed here applies Metropolis Gibbs sampling (with parameter rotations) to unknown geoacoustic and sound-speed profile parameters, whereas a two-dimensional variant of heat-bath GS is applied to source location to exploit the properties of normal mode modeling to provide efficient sampling over complicated $r-z$ surfaces. Analytic results are applied for unknown source amplitude/phase and data uncertainties which results in an efficient implicit formulation that does not require explicit unknown parameters to represent these quantities.

To quantify localization uncertainty, probability ambiguity surfaces and the integrated probability of localization within an acceptable region were considered for acoustic data measured at a shallow-water site in the Mediterranean Sea where previous geoacoustic inversions had been carried out. Localization uncertainty distributions were found to be strongly dependent on prior environmental information and on data quality and quantity. For single-frequency data at low SNR (-3 dB), high-probability localizations were obtained consistently only for small prior uncertainties on both seabed geoacoustic parameters and water-column sound-speed parameters. The particular case considered here appeared to be more sensitive to uncertainty in seabed properties than water-column properties. For single-frequency data at higher SNR ($+3$ dB) or for multifrequency data, high-probability localizations were obtained consistently even with wide prior uncertainties for environmental parameters. Hence, reliable localization requires a sufficient combination of prior (environmental) information and data information.

ACKNOWLEDGMENTS

[1] J. M. Ozard, "Matched field processing in shallow water for range, depth, and bearing determination: Results of experiment and simulation," J. Acoust. Soc. Am. **86**, 744–753 (1989).

[2] A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1993).

[3] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).

[4] A. Tolstoy, "Sensitivity of matched field processing to sound-speed prone mismatch for vertical arrays in a deep water Pacific environment," J. Acoust. Soc. Am. **85**, 2394–2404 (1989).

[5] M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **92**, 2770–2783 (1992).

[6] S. E. Dosso, M. L. Yeremey, J. M. Ozard, and N. R. Chapman, "Estimation of ocean-bottom properties by matched-field inversion of acoustic field data," IEEE J. Ocean. Eng. **18**, 232–239 (1993).

[7] P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**, 770–782 (1994).

[8] M. D. Collins and L. Fishman, "Efficient navigation of parameter landscapes," J. Acoust. Soc. Am. **98**, 1637–1644 (1995).

[9] P. Gerstoft, "Inversion of acoustic data using a combination of genetic algorithms and the Gauss-Newton approach," J. Acoust. Soc. Am. **97**, 2181–2190 (1995).

[10] S. E. Dosso, M. J. Wilmut, and A. L. Lapinski, "An adaptive hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).

[11] P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).

[12] S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).

[13] S. E. Dosso and P. L. Nielsen, "Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data," J. Acoust. Soc. Am. **111**, 143–159 (2002).

[14] S. E. Dosso, P. L. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 208–219 (2006).

[15] M. Snellen, D. G. Simons, M. Siderius, J. Sellschoop, and P. L. Nielsen, "An evaluation of the accuracy of shallow water matched field inversion results," J. Acoust. Soc. Am. **109**, 514–527 (2001).

[16] M. Siderius, P. L. Nielsen, J. Sellschoop, M. Snellen, and D. Simons, "Experimental study of geoacoustic inversion uncertainty due to ocean sound-speed fluctuations," J. Acoust. Soc. Am. **110**, 769–781 (2001).

[17] C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **6**, 1–28 (2000).

[18] S. E. Dosso, "Environmental uncertainty in ocean acoustic inversion," Inverse Probl. **19**, 419–431 (2003).

[19] M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," J. Acoust. Soc. Am. **90**, 1410–1422 (1991).

[20] A. M. Richardson and L. W. Nolte, "*A posteriori* probability source localization in an uncertain sound speed, deep ocean," J. Acoust. Soc. Am. **89**, 2280–2284 (1991).

[21] J. A. Shorey, L. W. Nolte, and J. L. Krolik, "Computationally efficient Monte Carlo estimation algorithms for matched field processing in uncertain ocean environments," J. Comput. Acoust. **2**, 285–314 (1994).

[22] J. A. Shorey and L. W. Nolte, "Wideband optimal *a posteriori*, probability source localization in an uncertain shallow ocean environment," J. Acoust. Soc. Am. **103**, 355–361 (1998).

[23] S. L. Tantum and L. W. Nolte, "Tracking and localizing a moving source in an uncertain shallow water environment," J. Acoust. Soc. Am. **103**, 362–373 (1998).

[24] M. K. Sen and P. L. Stoffa, *Global Optimization Methods in Geophysical Inversion* (Elsevier, Amsterdam, The Netherlands, 1995).

[25] P. L. Nielsen, F. Bini-Verona, and F. B. Jensen, "Environmental and acoustic data collected south of the island of Elba during the PROSIM'97 experiment," SACLANT-CEN Document No. SM-357, SACLANT Undersea Research Centre, La Spezia, Italy (1999).

[26] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, The Netherlands, 1987).

[27] W. R. Gilks, S. Richardson, and G. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Champan and Hall, London, 1996).

[28] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustic* (AIP Press, New York, 1994).

[29] E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto-elastic ocean environments," J. Acoust. Soc. Am. **100**, 3631–3645 (1996).

Acoustic propagation under tidally driven, stratified flow

Steven Finette[a)] and Roger Oba
Acoustics Division, Naval Research Laboratory, Washington, DC 20375

Colin Shen and Thomas Evans
Remote Sensing Division, Naval Research Laboratory, Washington, DC 20375

Amplitude and phase variability in acoustic fields are simulated within a canonical shelf-break ocean environment using sound speed distributions computed from hydrodynamics. The submesoscale description of the space and time varying environment is physically consistent with tidal forcing of stratified flows over variable bathymetry and includes the generation, evolution and propagation of internal tides and solibores. For selected time periods, two-dimensional acoustic transmission examples are presented for which signal gain degradation is computed between 200 and 500 Hz on vertical arrays positioned both on the shelf and beyond the shelf break. Decorrelation of the field is dominated by the phase contribution and occurs over 2–3 min, with significant recorrelation often noted for selected frequency subbands. Detection range is also determined in this frequency band. Azimuth-time variations in the acoustic field are illustrated for 100 Hz sources by extending the acoustic simulations to three spatial dimensions. The azimuthal and temporal structure of both the depth-averaged transmission loss and temporal correlation of the acoustic fields under different environmental conditions are considered. Depth-averaged transmission loss varies up to 4 dB, depending on a combination of source depth, location relative to the slope and tidally induced volumetric changes in the sound speed distribution.
[DOI: 10.1121/1.2713724]

I. INTRODUCTION

A significant issue in sonar system evaluation is the prediction of array performance in littoral regions where significant ocean variability is present. In this respect, the ability to simulate sensor array performance in controlled but realistic synthetic ocean environments can represent a significant contribution to a performance prediction scheme. Realistic simulation of acoustic propagation in littoral regions depends on knowledge of several quantities that characterize the ocean waveguide structure: a volumetric sound speed distribution, a set of bottom parameters and waveguide boundary conditions. This information is collectively referred to as an "environmental model." As a step toward developing a predictive simulation capability, this paper describes a model of acoustic propagation in a littoral environment where submesoscale sound speed variability is computed in a physically consistent manner with respect to fluid motion over variable bathymetry. For littoral regions such as continental shelf-break areas, the sound speed distribution has significant space-time dependence and, therefore, an environmental model of such a region should incorporate this variability in a consistent manner with respect to external forcing and internal flows. Because of the quality of currently available numerical codes describing ocean acoustic propagation for a given sound speed distribution, the ability of an environmental model to faithfully represent a natural oceanographic environment determines, to a large extent, the predictive capa-

bility of an acoustic modeling effort. The emphasis here is on the numerical model and several example computations are presented. A direct comparison of numerical results with experimental data for a particular ocean environment is beyond the scope of this paper; that issue will be addressed in future work in order to assess the predictive capability of this approach.

A specific case is considered for illustrative purposes, that of an idealized continental shelf-break environment where tidally driven flow of stratified water over variable bathymetry induces volume variability in the form of internal tides and nonlinear wave packets. The latter are often referred to as solitary wave packets or solibores. These oceanographic features impact acoustic field propagation and subsequent array performance through their effect on the sound speed field. The sound speed distribution is determined here by first solving the Navier–Stokes equations of fluid dynamics which describe the space–time temperature, salinity and non-hydrostatic pressure fields, and then mapping these distributions into sound speed using a phenomenological relationship. Simulations of the influence of tidal flow on acoustic propagation are presented using this environmental model in conjunction with acoustic field calculations provided by two-dimensional (2D) and three-dimensional (3D) wide-angle parabolic equation codes.

A combined oceanographic-acoustic experiment, SWARM95,[1] served as a stimulus for a number of modeling efforts related to the interaction of acoustic fields with both internal tides and solitary wave packets. Although these studies used relatively simple models to describe the volume variability of the sound speed field linked to these oceano-

---

[a)]Electronic mail: steven.finette@nrl.navy.mil

graphic features, they yielded significant insights into the nature of this interaction. The early emphasis was first placed on scintillation index[2] and mode coupling[2–5] along the direction of propagation of the wave packet using simplified packet wave-form shapes to describe the isopycnal displacements and resulting sound speed perturbations. More complex waveforms based on weakly nonlinear theories of solitary wave propagation were also considered. These included the cnoidal[6] and dnoidal[5,7] elliptic functions that more closely resembled some of the isopycnal variations found in experimental data. The relationship between the acoustic field structure and source/receiver orientation with respect to the solitary wave propagation vector was described using $N \times 2D$ single frequency[6] and pulse[7] calculations of the acoustic field, without the inclusion of horizontal refraction. The effects of horizontal refraction on both acoustic field intensity and horizontal (cross range or transverse) coherence are inherently 3D and have been analyzed using both analytic[3,8,9] and numerical simulation techniques.[10,11] Relationships between horizontal array beamforming and horizontal refraction induced by solitary wave packets have been described recently.[12]

Mesoscale ocean models have been applied to describe the evolution of large-scale sound speed distributions for some time.[13,14] This class of ocean models filter out spatially localized internal gravity waves that are ubiquitous in continental shelf regions. As these waves have a significant effect on the amplitude and phase of an acoustic field propagating through the region, a more comprehensive hydrodynamic model of littoral oceanography is considered here by describing the environment with a submesoscale ocean model, representing recent developments in simulating ocean dynamics.[15–19] The submesoscale models can naturally include, in a physically consistent formalism, the generation, propagation, and dissipation of internal waves driven by tidal forcing as well as other oceanographic features such as fronts, jets and overturning bores. Recently, there have been attempts to evaluate the influence of some of these finer scale ocean structures on acoustic propagation.[20–22] Our view is that the extra effort involved with the design and implementation of submesoscale ocean models to estimate the sound speed field for use in acoustic propagation studies of sonar performance is mitigated by a corresponding increased flexibility in describing littoral environments, the potential enhancement to acoustic predictive schemes and future advances in computer hardware. The issue of whether these models enhance predictability over that obtainable through either mesoscale modeling or by use of approximate approaches involving, e.g., weakly nonlinear Korteweg–de Vries (KdV) descriptions of internal waves is an important issue in performance prediction but is not considered here. It can only be resolved by quantitative comparisons of simulation results with experimental data, taking into account the environmental uncertainty in the parameters, fields, and model initialization describing the simulated environment. Such model/data comparisons have a host of issues that would need to be considered, though it is worth making a brief comment here to distinguish between the verification and validation procedures needed to develop a simulation-based prediction scheme. The results presented in this paper are based on codes that were benchmarked as part of a validation procedure to ensure that they properly represent, in a numerical form to some level of accuracy, the mathematical description of the system (ocean environment and acoustic field). Verification, on the other hand, depends on computing metrics which assess the ability of the numerical model to predict or represent experimental acoustic data. The spatial and temporal sampling requirements needed to execute a submesoscale code reflect the underlying scales of the physics to be represented and correspond to much finer sampling of the boundary/initial conditions and ocean volume than is normally available in partially controlled, at-sea experiments for which temperature and salinity are sparsely sampled. The resulting information deficit presents a significant challenge in the validation phase where the "value-added" of submesoscale modeling over other approaches for making acoustic predictions needs to be addressed.

Section II briefly describes the environmental model and implementation, whereas Sec. III demonstrates its use through examples of acoustic propagation in a canonical shelf-break environment where it is assumed that the sound speed dynamics is driven exclusively by tidal flow. Both signal gain degradation and detection range are computed under stratified flow as a function of frequency and related to the oceanographic state of the environment as a function of time. The azimuthal dependence of the acoustic field on both bathymetry and tidal flow is then discussed in the context of a simple cross-range extension of the shelf break environment in conjunction with 3D acoustic simulations. A summary and discussion are given in Sec. IV.

II. ENVIRONMENTAL DYNAMICS

The submesoscale hydrodynamic model used in the following analysis is briefly considered in this section.[17,18] The ocean environment is described through the equations of motion for an incompressible fluid in the Boussinesq approximation, where the density variations are retained only in the buoyancy term.[23] This approximation is excellent for littoral ocean environments. The dynamics are represented by a set of coupled partial differential equations describing the Lagrangian fluid particle velocity $\boldsymbol{u}(\boldsymbol{r},t)$, pressure $p(\boldsymbol{r},t)$, density $\rho(\boldsymbol{r},t)$, temperature $T(\boldsymbol{r},t)$, and salinity $S(\boldsymbol{r},t)$ fields, where $\boldsymbol{r}=(x,y,z)$, and $\boldsymbol{u}=(u,v,w)$ specify the three Cartesian velocity components $(dx/dt,dy/dt,dz/dt)$. Water depth is associated with the z direction, negative toward increasing water depth. Here, the ocean surface is free to move under external tidal forcing, a feature often suppressed in ocean modeling. The momentum equation is expressed as

$$\frac{D\boldsymbol{u}}{Dt} + \boldsymbol{f} \times \boldsymbol{u} = -\frac{\nabla p}{\rho_o} + \boldsymbol{g}\frac{\rho}{\rho_o} + \mu\nabla^2\boldsymbol{u} \qquad (1)$$

Tidal forcing is applied at the open boundary and is described later in this section. Other environmental forcing such as that associated with wind stress could be included but are not considered here. A viscous term, proportional to viscosity, μ, is included for numerical stability and may be space dependent. The equilibrium density is given by ρ_o,
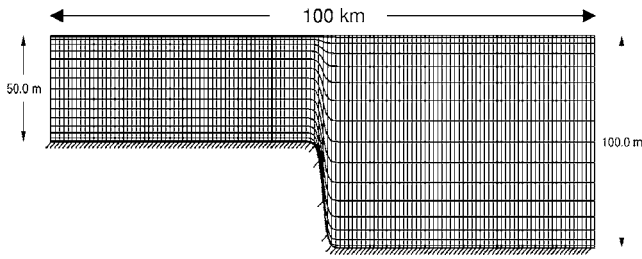
FIG. 1. Cross-range view of the terrain following coordinate system used to simulate a canonical shelf-break ocean environment. The acoustic field is computed over a 30 km range centered on the shelf break.

whereas the vector $\boldsymbol{f} \times \boldsymbol{u}$ represents Coriolis acceleration. The advective derivative operator is defined by $D/Dt \equiv \partial/\partial t + \boldsymbol{u} \cdot \nabla$. The second term in this operator represents nonlinear interactions and is responsible for the generation of internal tides and solibores arising from tidal forcing. For incompressible fluids, the equation of continuity becomes

$$\nabla \cdot \boldsymbol{u} = 0. \tag{2}$$

Temperature and salinity fields are described by a pair of advection/diffusion equations, where

$$\frac{D\{T,S\}}{Dt} = \tau_{\{T,S\}} \nabla^2 (T,S) \tag{3}$$

with the diffusivity defined by $\tau_{\{T,S\}}$; the particular tracer is denoted by the subscript. The density field, ρ, is a function of space and time and it is related to temperature, salinity, and pressure distributions by an equation of state

$$\rho = \rho(T,S,p). \tag{4}$$

The IES-80 equation of state[24] is incorporated into this model to explicitly define Eq. (4). Together with a set of boundary and initial conditions, Eqs. (1)–(4) completely specify the fluid flow. The sound speed field $c(\boldsymbol{r},t)$ is a derived quantity, obtained from a phenomenological relationship[25] involving T, S, and p once the previous set of equations are solved as a function of position for each time step. For computational purposes, the previous equations are mapped into an equivalent formulation treating the vorticity $\boldsymbol{\xi} = \nabla \times \boldsymbol{u}$ rather than the particle velocity.

For the example presented in the next section, 2D hydrodynamic computations are performed over a range of 100 km in water of variable depth, between a value of 50 m on the shelf and 100 m beyond the shelf break. The oceanographic grid represents a terrain following coordinate system and is shown in Fig. 1. Figure 1 represents a canonical shelf-break environment with a maximum slope of 2.4°; the bathymetry is modeled in terms of a hyperbolic tangent function. The range step was fixed at 25 m, with the number of depth steps fixed at 65 for each range step. Because of the variable bathymetry and the cosine grid in the vertical, the depth resolution was dependent on range. Tidal forcing is implemented by applying a sinusoidal surface displacement at the right-hand boundary, with an amplitude of 2 m and a period of 12.4 h, with the latter corresponding to an M2 tidal cycle. The boundary conditions imposed on the environmental model are as follows. The model domain is closed on the

coastal (shallow) side, $x=0$, by a vertical wall, where the boundary conditions are $u=0$, $\partial v/\partial x=0$, $\partial w/\partial x=0$, and $\partial \eta/\partial x=0$ with η representing the surface displacement. The boundary conditions on the open ocean side ($x=100$ km) are, in addition to the tidal forcing given earlier, specified as $\partial v/\partial x=0$ and $\partial w/\partial x=0$. No condition is imposed on u at the open ocean boundary. The vertical boundary conditions are $\partial u/\partial z=0$, $\partial v/\partial z=0$, and $\partial \rho/\partial z=0$ at the free ocean surface and bottom. For acoustic computations, a 30 km subset of this grid was chosen and centered on the shelf break. The sound speed field was interpolated to a rectangular grid for each time step prior to performing acoustic simulations.

III. SIMULATION RESULTS FOR A SHELF-BREAK ENVIRONMENT

The shelf-break environment was chosen as an example to illustrate some features of acoustic field variability caused by tidally forced stratified flow over variable bathymetry. Some previously measured data is used for environmental initialization, but no effort is made to accurately model a specific shelf break region. For each environmental snapshot computed by the approach discussed in Sec. II, the sound speed distribution is processed through a wide-angle parabolic equation code incorporating a split-step Padé solver.[26] Assuming the environment is frozen for each snapshot, the acoustic field between 200–500 Hz is computed as a function of source depth and frequency over a 30 km range subset of the full 100 km environmental simulation, with the acoustic computational window centered at the point of maximum slope on the shelf break (see Fig. 1). Complex acoustic fields are then "stored" on vertical arrays located at selected ranges along this 30 km track.

A tacit assumption concerning the environment has been introduced here. For computational reasons we are effectively ignoring along-slope (i.e out of plane) oceanographic variability. For many shelf-break regions, significant along-slope variations in both bathymetry and flow properties exist. Although fully 3D hydrodynamic computations of the environment and their effect on out-of-plane acoustic propagation are beyond the scope of this paper, we do consider a simplified scenario in Sec. III B 2 where azimuthal variations in the acoustic field are introduced in the following way. The 2D environmental snapshots are extended in the along-slope direction, mimicking a 3D shelf-break without along-slope oceanographic or bathymetric variability. The environment is therefore chosen to be invariant with respect to the along-slope direction, but the acoustic field can exhibit a dependence on azimuth. A wide-angle 3D parabolic equation[27] is then used to evaluate the azimuthal and time dependence of the acoustic field on vertical arrays located at a fixed radial distance of 10 km from acoustic sources placed at several locations near or in the shelf break region. The source frequency in this case is 100 Hz.

A. Sound speed variations caused by tidal forcing

Several environmental snapshots of the time-evolving sound speed environment and resulting acoustic transmission losses within the 30 km computation window are shown in
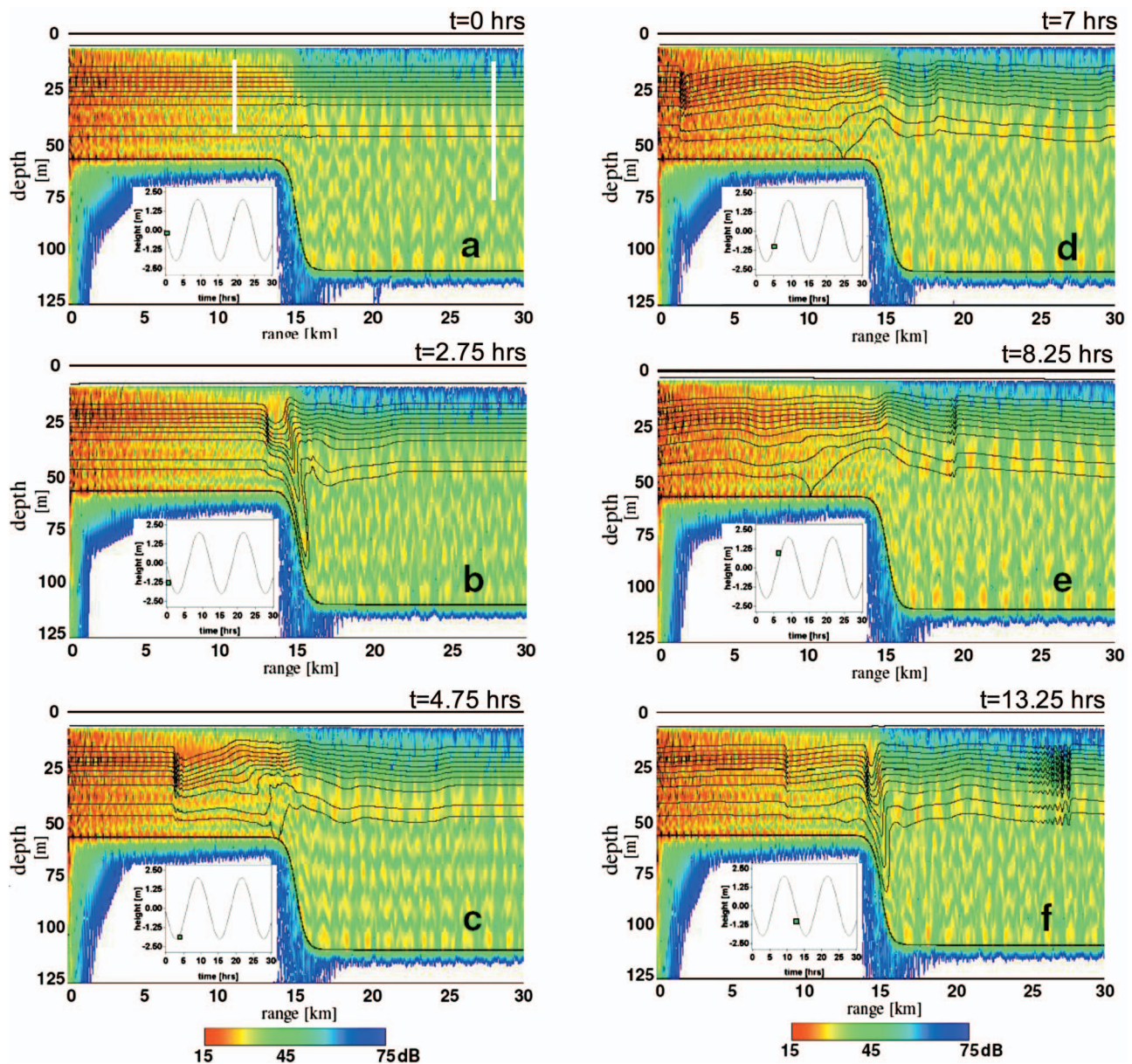
FIG. 2. Sequential set of environmental snapshots and associated transmission loss obtained over a period of approximately 13 h. For each, the inset shows the corresponding phase of the tidal cycle during which the sound speed field was computed. Sound speed is represented by the black contour lines, and initially ($t=0$) corresponds to a range-independent summer profile. A thick black line at the top of each figure represents the maximum water height. The color overlays describe transmission loss at 400 Hz for a source at 15 m depth, located at the left edge of each picture. Geometric spreading is not included in the plots.

Fig. 2, which covers a 30 h time span [approximately 2.5 semidiurnal (M2) tidal cycles]. The environment in Fig. 2(a) corresponds to the starting sound speed distribution for the simulation, computed from the initial temperature and salinity distributions. The initial sound speed field is chosen to be range independent over the 30 km propagation track at time $t=0$ h. The depth dependence of the sound speed represents a typical summer thermocline for the New Jersey shelf,[1] with a mixed layer of about 10 m below the surface and sound speed variations of approximately one percent from the surface to the bottom. The range independence of the sound speed field is not typical for this region; it is introduced here for simplicity in order to clearly illustrate the generation of internal tides and solibores without the additional complica-

tion associated with the existence of a range- and depth-dependent cold water pool that can extend out to the shelf break.

The fluid velocity is initialized to zero, corresponding to a quiescent ocean referenced to time $t=0$ h. Tidal forcing is initialized as shown in the inset for Fig. 2(a), so that during spin-up the conditions corresponded to ebb flow. Choosing on-shore flow during spin-up results in a different transient sound speed response that eventually settles into a temporal sequence of environments similar to that described here. The thin black lines illustrated in the water column are sound speed contours, whereas the color coded overlays represent the acoustic transmission loss for a 400 Hz source located at the range origin ($r=0$) and source depth of 15 m. The source

is fixed relative to the free ocean surface; this is equivalent to a configuration where the source is moored to a surface ship. Geometric spreading has not been included in Fig. 2. Bottom parameters (sound speed, density, and attenuation) are chosen to be average values typically found in the New Jersey Shelf, though range dependence in these parameters is ignored. Note that our purpose here is to illustrate the use of the model in a relatively simplified setting using reasonable parameter values, rather than to perform a detailed analysis of a specific environment. An analysis of specific environments will be pursued in future efforts. White vertical lines drawn in Fig. 2(a) locate the positions of the two vertical arrays which are used to process the simulated field data along a 2D range/depth slice. Array positions are fixed at 11 and 28 km from the source location. In Sec. III B 2, an alternative set of arrays are positioned at fixed range and variable azimuth with respect to sources placed near or within the slope-break region in order to investigate azimuthal field dependence in this environment.

As the barotropic tide continuously forces stratified water over the shelf edge, a significant vertical component of the buoyancy force develops at this location and acts as a local wave generation mechanism in which energy is radiated away from the shelf break. This energy is predominantly in the form of an internal tide, illustrated in Fig. 2(b), where an internal tide has recently formed and moved onto the shelf. It is located at a range of about 13 km from the source and is traveling shoreward. The sound speed contours indicate that the temperature has been advected downward at the shelf break during this time. A few hours later, Fig. 2(c) shows that the internal tide has progressed about 7 km shoreward and developed oscillations on the trailing edge, indicating a transformation of the internal tide to a solibore with a depth dependence dominated by the first internal wave mode. A few hours later, the solibore has continued to propagate toward the acoustic source and upwelling is observed over the shelf break [Fig. 2(d)]. As time progresses another internal tide has formed, begun to transform into a solibore at a range of about 20 km and moves seaward [Fig. 2(e)]. This solibore continues to add oscillations as it progresses seaward [Fig. 2(f)]. A solibore is also seen at a range of about 8 km traveling shoreward and another internal tide is beginning to form over the shelf break at a range of approximately 14 km. A detailed analysis of the ocean dynamics for a qualitatively similar environment is given in Ref. 15 in which a rigid lid, rather than a free surface model, is used to simulate internal wave generation. Changes in water depth caused by free surface movement in response to tidal forcing are viewed most clearly by comparing Figs. 2(b) and 2(e), where the thick black line at the top of each figure represents the upper limit of the water height. Note that tidal waves have wavelengths on the order of 10^2 km, the length of the computational domain. Over the 30 km range for the acoustic computational grid, the dominant effect of the tide is to raise or lower the water column height uniformly as a function of time. Small scale surface waves would cause spatially local water depth variations over short time intervals, but the parabolic equation code used for the acoustic calculations cannot handle those variations directly. An alternative implementa-

tion of the parabolic equation[28] could be used to include small scale surface waves, but the inclusion of the corresponding hydrodynamic component over the domain of interest would be computationally prohibitive.

B. Acoustic field variability caused by tidal forcing

1. Acoustic response along a range/depth slice

The transmission loss patterns in Fig. 2 are indicative of time dependence in multi-mode propagation. The patterns themselves are determined by several factors such as (1) the source depth and local environment at the source which determines the initial modal distribution of energy, (2) tidally induced modulation of water depth which can introduce or cut off a mode or modify the relative phasing between modes, and (3) volume variability associated with the internal tides and wave packets induced by tidal forcing. Since these factors are coupled together through tidal forcing, it is difficult to separate out the individual contributions to the overall temporal and spatial evolution of the transmission loss pattern.

In order to quantify variability in the acoustic field, the vertical arrays [cf. Fig. 2(a)] are used to process the simulated field data. They span water depths between 7.5 and 40 m at the shallow array site, 7.5–80 m at the deep site. Both the acoustic source and the arrays were assumed to be moored from surface ships; as the free surface moves in response to the tide, the source and array depths remain fixed relative to the surface but change relative to the fixed bottom. Two time windows, A and B, with durations of 30 min each are selected from the 30 h simulation. The complex pressure $P(r,z,t,\omega)$ at frequency $\omega/2\pi$ is recorded on each hydrophone every 20 s within a window, for a total of 90 complex acoustic field vectors per array for each window. The frequency band for the acoustic field computation was 200 $\leq \omega/2\pi \leq 500$ Hz, where acoustic computations are performed with a 1 Hz resolution across the band yielding a total of 54,000 field vectors. The starting time for window A corresponds to the environment in Fig. 2(d), whereas the starting time for window B corresponds to the environment in Fig. 2(f). Window A is representative of an environment where a solitary wave packet passes through the acoustic source position, changing the local sound speed distribution near the source, while window B describes a situation where a packet crosses the array located 28 km from the source and the local source environment is approximately time independent. In both cases, the sound speed field between the source and the array is space and time dependent. Within each window a function $\Gamma(\omega,t,t_0)$, defined below, is computed for $t_0 \leq t \leq t_0 + \Delta$ at frequency $\omega/2\pi$, where t_0 is the reference starting time for evolving the environment in window A or B, and Δ is the window width of 30 minutes.

Let $e(\omega,t_0)$ represent a "ground truth" vector of complex pressure fields, $[P(r,z_1,t_0,\omega)...P(r,z_N,t_0,\omega)]^T$ received on a vertical array of N phones at time t_0 during either window A or B, and let $f(\omega,t)$ represent a replica vector of complex pressure fields on the same array at time $t \geq t_0$. Then $\Gamma(\omega,t,t_0)$ is defined by a ratio of inner products given by

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Finette et al.: Propagation under stratified flow   2579

r = 11 km   source depth = 27 m

a

r = 11 km   source depth = 42 m

c

signal gain degradation [dB]
-25 -21 -17 -13 -8 -4 -0

r = 28 km   source depth = 27 m

b

r = 28 km   source depth = 42 m

d

signal gain degradation [dB]
-25 -21 -17 -13 -8 -4 -0
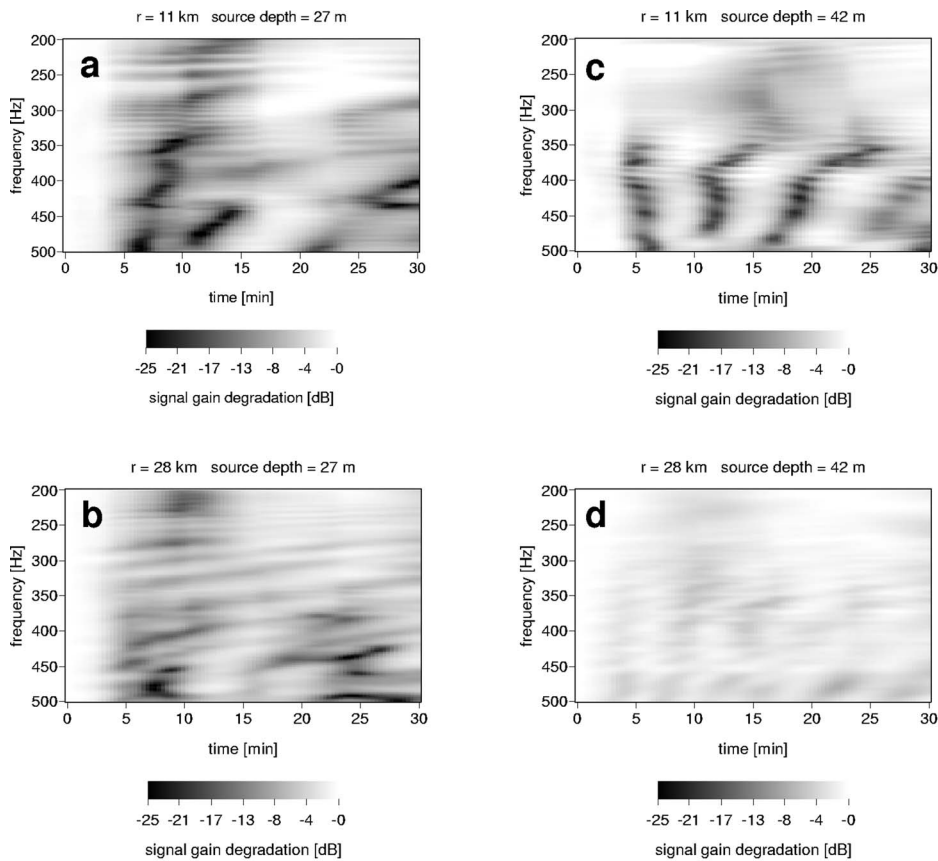
FIG. 3. Signal gain degradation computed from Eq. (5) as a function of frequency and time during window A, for two source depths (27 and 42 m) and two array locations (11 and 28 km). The degradation was computed with a free ocean surface.

$$\Gamma(\omega,t,t_0) = \frac{|e^* \cdot f|^2}{|e^* \cdot e|^2} = \frac{|f(\omega,t)|^2}{|e(\omega,t_0)|^2} \cdot |\hat{e}^*(\omega,t_0) \cdot \hat{f}(\omega,t)|^2 \quad (5)$$

with the caret denoting a unit vector and an asterisk representing the conjugate transpose. Equation (5) is useful for the assessment of array performance in a time-varying medium, and can be interpreted as a measure of signal gain degradation within the time window. It provides an estimate for the period in which a measured field could be used as a replica in a linear matched field processor before the peak power degrades significantly. The middle equality in Eq. (5) then represents the matched field processor power relative to the expected power. Algebraic manipulation yields the second line of Eq. (5), separating the total degradation associated with the full complex field into the product of its individual amplitude and phase contributions, respectively.[29] Signal gain degradation, in the form $10 \log \Gamma$, is discussed below for the important case of propagation under tidally driven stratified flow. We note here that in the next subsection a different interpretation is given for the vectors e and f, and Eq. (5) is then interpreted as a temporal correlation coefficient measured across the array aperture.

Several plots of signal gain degradation (SGD) are given in Fig. 3 for results obtained by propagation through the 30 min sequence of environments associated with window A. The images are presented as functions of time and source frequency (200–500 Hz band) at two source depths (27 and 42 m) for each of the two arrays. The frequency-time degradation patterns reveal significant complexity, even for this relatively simple environment. Representing the full degradation (i.e. a combination of amplitude and phase contribu-

tions), they show that over the first 2 to 3 min the SGD is less than about 2 dB while for times greater than about 3 min, the SGD is usually (but not always) significantly larger with considerable variation as a function of frequency and time. The increased degradation after a few minutes implies mismatch (i.e., temporal decorrelation) between the signals e and f. Temporal correlation of narrow-band signals that fall within this frequency band has been measured[30,31] and modeled[5,32,33] in shallow water environments exhibiting significant internal wave activity, and full field correlation times on the order of two or three minutes are typical. The correlation times determined in those studies were found to be mode dependent, with the first mode tending to have the longer correlation time since its major contribution to the field as a function of depth spans a vertical region below the thermocline and is less susceptible to sound speed perturbations in the upper part of the water column.[5] The measured values are consistent with the results shown in Fig. 3. For acoustic propagation in the 200–300 Hz band, Fig. 3(a) shows a clear recorrelation of the field after about 15 min. To a somewhat lesser extent, a similar effect is seen in Figs. 3(b) and 3(d) for different frequency bands and time slices within the window. Within the 350–500 Hz band, Fig. 3(c) displays a comparatively stronger recorrelation and also shows a temporally periodic degradation structure. Partial recorrelation of the field has been noted in the SWARM95 data and interpreted as a periodic feature occurring when the relative phase retardation between dominant mode pairs approaches 2π radians.[33] With the exception of Fig. 3(d), these plots indicate that the degradation is enhanced as the packet passes

a r = 11 km   source depth = 27 m

c r = 11 km   source depth = 42 m

-25 -21 -17 -13 -8 -4 -0
signal gain degradation [dB]

b r = 28 km   source depth = 27 m

d r = 28 km   source depth = 42 m

-25 -21 -17 -13 -8 -4 -0
signal gain degradation [dB]
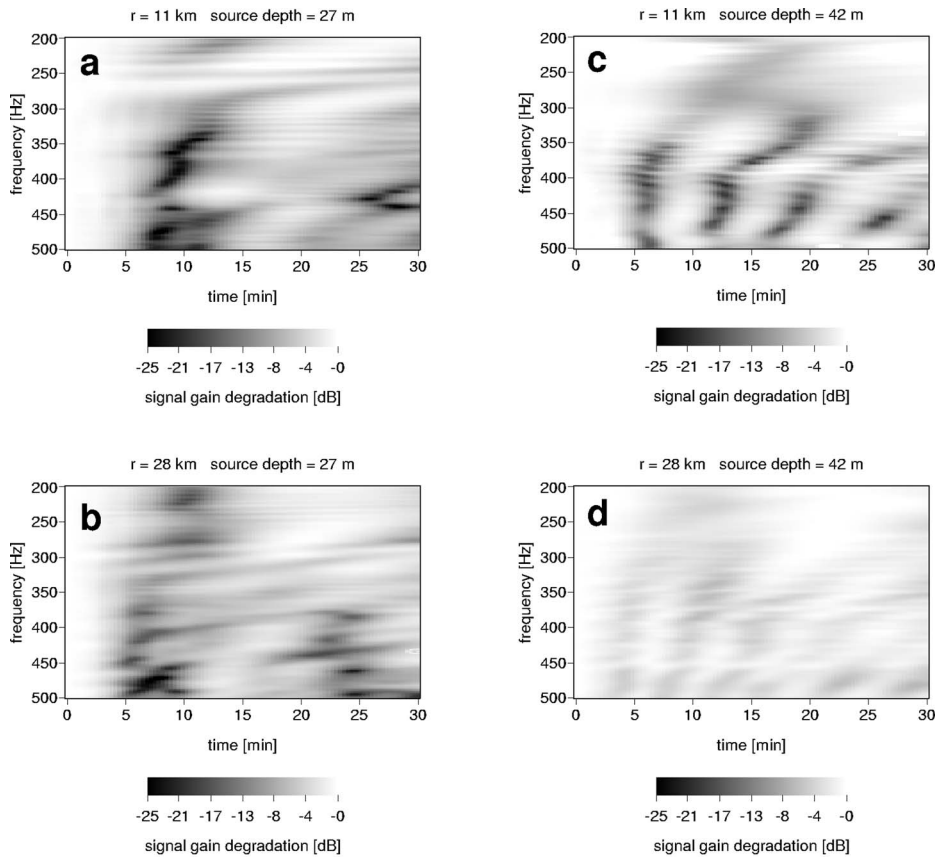
FIG. 4. Signal gain degradation computed as in Fig. 3, but with a fixed ocean surface.

over the source position, a result discussed by several authors[3,5] using simple internal wave models. The result here is more complex and difficult to interpret as the environment changes along the entire propagation path, whether a solitary wave packet is present or absent.

Results from a number of studies support the view that the frequency and time dependence of the degradation illustrated in Fig. 3 is likely caused by a combination of several time dependent effects linked to tidal flow: mode coupling along the acoustic track,[2–5] water depth variation[34,35] and redistribution of modal energy at the source.[2–5] The first effect is an indirect consequence of the stratified flow over variable bathymetry, which induces range and time-dependent internal wave perturbations of the mean sound speed field. While a conclusive analysis using modal decomposition is beyond the scope of this paper, previous work has shown that mode coupling is a function of both acoustic frequency and the range gradient of the resulting sound speed perturbations. It alters both the particular modes received at the array and their relative phase relationships,[2–5,7] so that mismatch between temporally separated signal transmissions is expected to occur. Water depth variations due to tidal flow can modify the mode distribution and phasing between modes, possibly by introducing a new mode or eliminating a mode that falls below cutoff. Over the particular short duration time windows considered in this example, water depth changes due to the tide have a small effect on SGD as discussed below, but over a longer time scale (hours) it can be more significant.[34,35] The third effect is that the local sound speed distribution in the immediate vicinity of the source changes due to sound speed perturbations passing through the region,

and this local environment determines the mode distribution excited by the source.[2,3,5] This distribution, in conjunction with bathymetric variation and mode coupling in the water column, can play a significant role in determining the energy distribution at the receiving arrays.

An attempt is made here to assess the relative importance on SGD of individual contributions of both tidally induced water depth variations and internal wave induced sound speed perturbations. A separate set of acoustic simulations are performed in the environment described previously, except that for these field computations the water depth is held fixed throughout the time window using the water depth present at time t_0. This case represents a situation where tidally induced water depth changes are ignored but volumetric sound speed perturbations caused by the tide are included. Examples of signal gain degradation under this constraint are shown in Fig. 4 for time window A. Comparing Figs. 3 and 4 one can detect small differences in degradation that can be attributed to changes in water depth. These depth variations are quite small, approximately 0.3 m over each of the 30 min time periods, and weakly alter the temporal variation of SGD in selected frequency bands. It is clear, for example, that most of the roughly periodic structure of the degradation above 350 Hz in Fig. 3(c) is primarily, though not exclusively, due to volume variability since the corresponding plot without surface variation [Fig. 4(c)] shows similar behavior. The effect on degradation of changing water depth is more significant for propagation on the shelf for the shallow source depth [Fig. 4(a)], probably due to the fact that the percentage change in water depth is about twice that found at the array located 28 km from the source.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Finette *et al.*: Propagation under stratified flow   2581
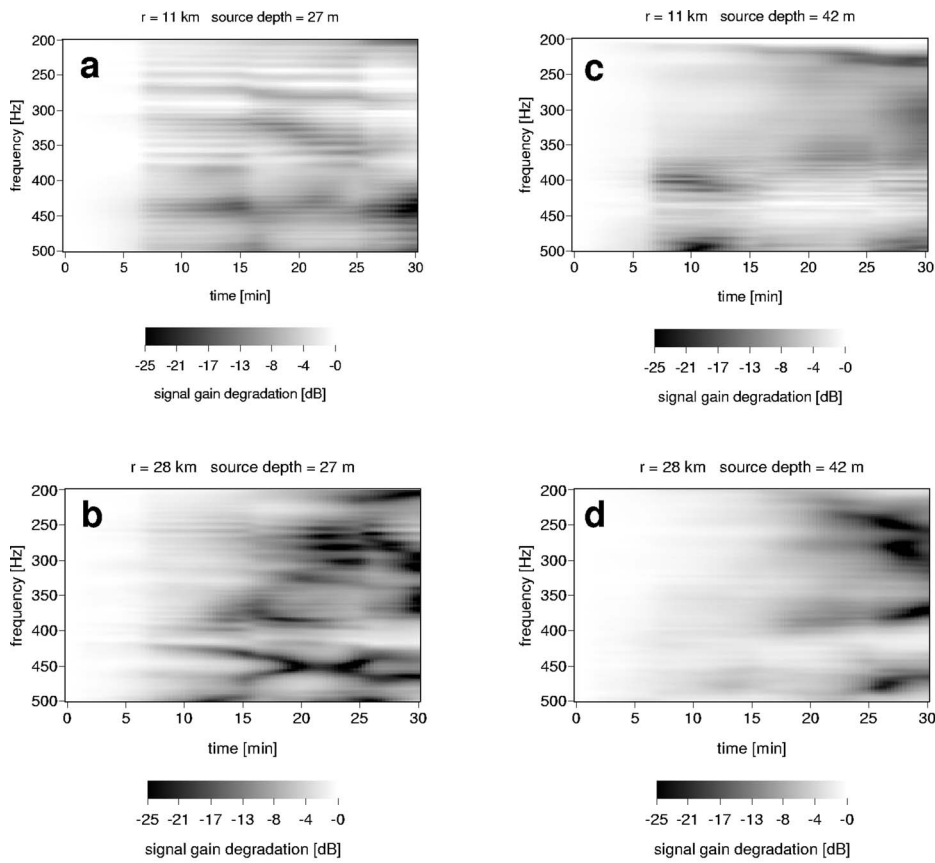
FIG. 5. Signal gain degradation for window B, computed with a free ocean surface.
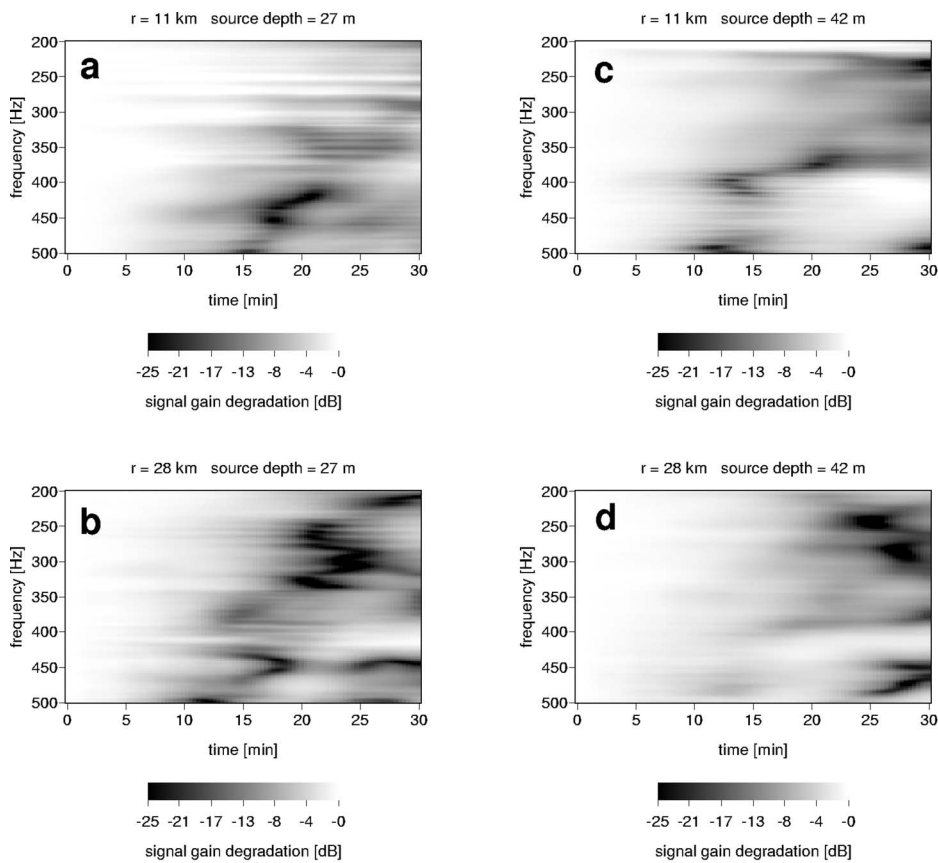
FIG. 6. Signal gain degradation for window B, computed with a fixed ocean surface.

Finette *et al.*: Propagation under stratified flow
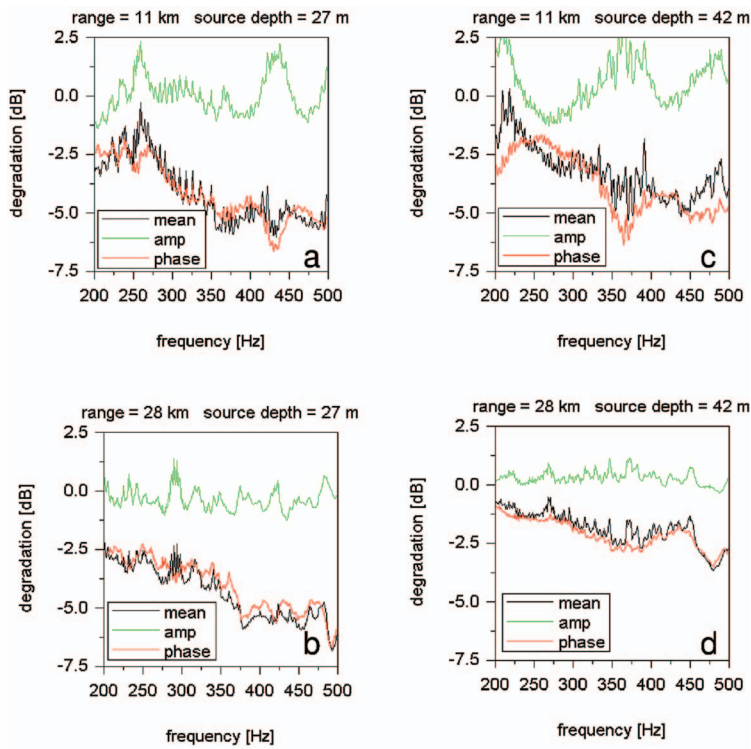
FIG. 7. Decomposition of signal gain degradation into amplitude and phase contributions for window A, computed with a free ocean surface.

Similar comments can be made for both the (full field) SGD plots shown in Fig. 5, which correspond to the evolving ocean environment in window B, and the associated fixed water depth case illustrated in Fig. 6. Again, there are small differences between the corresponding fixed surface/free surface degradation plots for most time-frequency regions, indicating the SGD is primarily caused by volume variability for short time windows. As was the case for the sequence of snapshots in window A, the most significant acoustic changes occur for the array located on the shelf. The environmental conditions along the track are quite different for window B relative to those in window A. For window B there is significant downwelling [Fig. 2(f)] that is not present in window A [Fig. 2(d)]. In addition, a large solitary wave packet crosses the receiving array during time window B. The full field degradation in Figs. 5(c) and 6(c) are strikingly different from the corresponding SGD computed with the environment in window A. Note that because the environ-
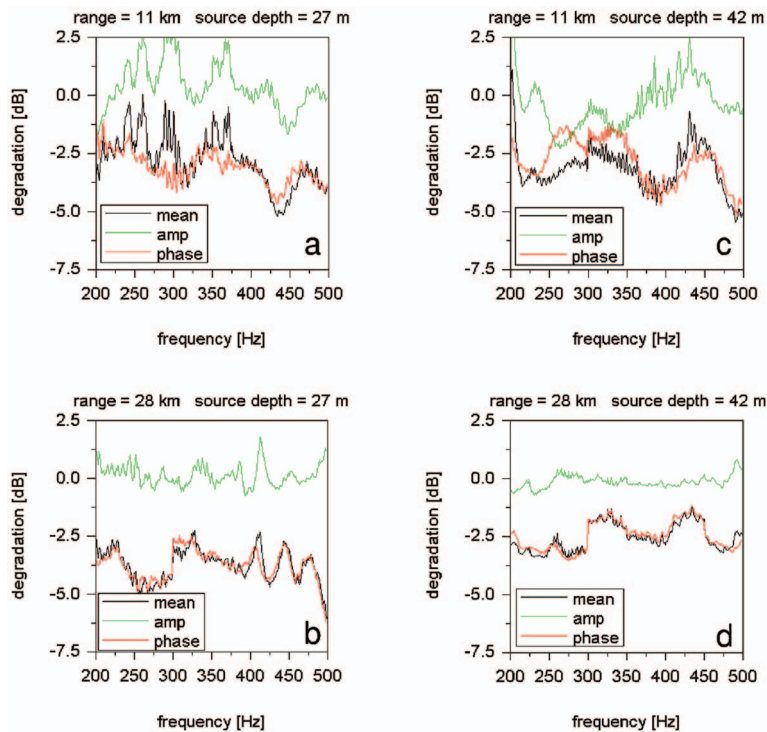
FIG. 8. Decomposition of signal gain degradation into amplitude and phase contributions for window B, computed with a free ocean surface.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Finette *et al.*: Propagation under stratified flow   2583

mental dynamics is nonstationary, some caution should be applied in extrapolating these results to arbitrary time windows during different phases of the M2 tidal cycle.

It is well known that phase fluctuations dominate variability in both transmission loss and field coherence. For example, measurements of array gain degradation for data acquired on the New Jersey shelf are consistent with the prediction that phase contributions dominate the degradation.[29] Results illustrated in Figs. 3–6 are consistent with phase dominance, a conclusion drawn from a direct comparison of frequency-time SGD plots for the individual amplitude and phase contributions obtained through Eq. (5). Instead of presenting those figures, this result is illustrated here in an alternative format by averaging the individual contributions to degradation over the time windows for each acoustic frequency. The results are shown in Fig. 7 corresponding to window A and Fig. 8 corresponding to window B. The black curve in each of these plots, labeled "mean," describes the full field calculations. For most frequency components, the trend is that the (full field) mean degradation follows the phase contribution. The trend is more closely followed for the longer range transmission (28 km), likely due to the relative importance of the gradual accumulation of phase shifts that occur as the acoustic field passes through small perturbations in sound speed over a longer propagation path. Similar trends are seen for the fixed surface cases (not shown), implying that the phase degradation is primarily due to volume variability.

An example is now presented which demonstrates how a representative sonar problem, that of detection range estimation, is affected by both tidal height variation and internal waves. Sound from a source located at 50 m depth and 15 km range seaward from the break is propagated directly up slope to receivers sampling the field at ranges between 20 and 30 km from the source. Detection range, r_d, based on a figure of merit FOM=66 dB is computed as follows. For the purpose of smoothing the data, a set of depths, ranges, and frequencies are included. The acoustic intensity is computed and averaged first over a 50 Hz bandwidth (in 1 Hz increments) starting at f_0, and then averaged over a 10 m square (20–30 m in depth and 10 m in range starting at r_0). Explicitly, the mean intensity is defined by

$$\bar{I} = \int_{r_0}^{r_0+10} \int_{20}^{30} \int_{f_0}^{f_0+50} \frac{I\,df\,dz\,dr}{10^2 \times 50}. \tag{6}$$

The spatially and frequency averaged TL is then given by $\overline{TL}=10\log_{10}\bar{I}$ for each time t in hours. The detection range is the range at which the TL crosses the FOM. However because $\overline{TL}$ is not strictly decreasing, $\overline{TL}$=FOM at several ranges, $r_{d,n=1,2,...}$ so that the detection range $r_d(t)$ is defined here as the average of the $r_{d,n=1,2,...}$ at time t. This also permits the use of the standard deviation $s_d(t)=\sigma(r_{d,n})$ to measure the spread of detection ranges. Finally, the time correlation gives a quantitative measure of how two time series vary simultaneously, for example, $\langle h_t, r_d \rangle = \int_0^{t_{max}} \tilde{h}_t(t)\tilde{r}_d(t)dt$ where the tilde indicates that the tidal height, h_t, and r_d are both zero mean and normalized. Because of the tidal variation, the bathymetry $b(r,t)=h_t(t)$
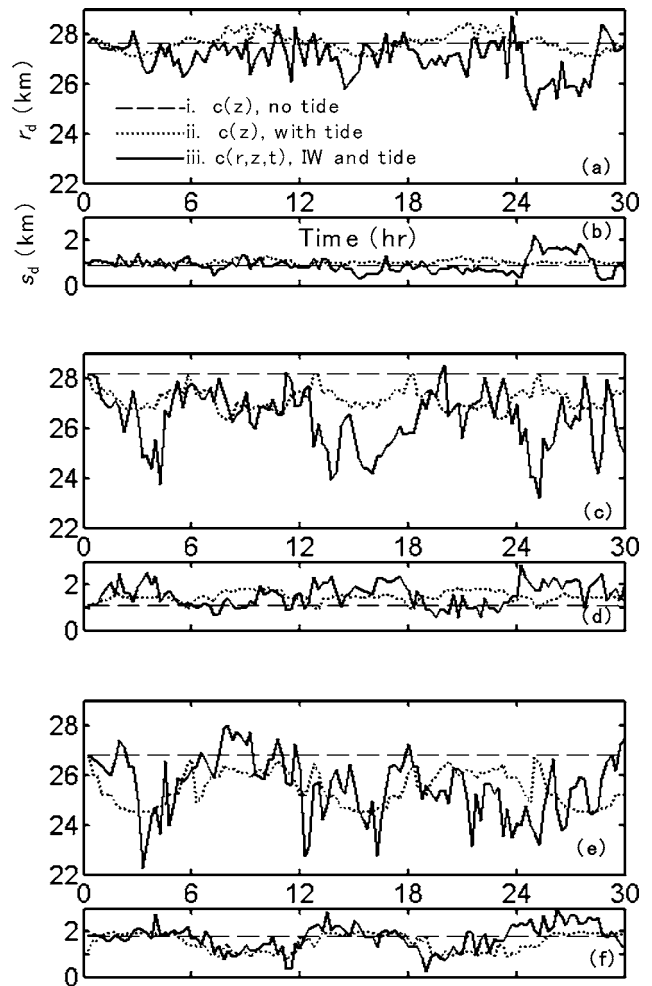
FIG. 9. Detection range estimates, r_d, and detection range variances, s_d, obtained over a 30 h time window as functions of acoustic frequency. (a,b) r_d, s_d for the 200–250 Hz band, respectively; (c,d) same quantities computed for the 300–350 Hz band; and (e,f) same quantities computed for the 400–450 Hz band.

$+b_0(r)$, where b_0 is the bathymetry at $t=0$. Figure 9(a) shows r_d for the 200–250 Hz band corresponding to three different cases: (i) the perfectly stratified sound speed distribution $c(z)$ without tidal variation (dashed line corresponding to the environment at $t=0$ h), (ii) the same sound speed profile, $c(z)$, but including only tidal variation h_t (dotted line), and (iii) inclusion of both internal waves and tidal variability $c(r,z,t)$ (solid line). In case (ii) $c(z)$ is required in the interval $0 \le z \le b(r,t)$ for each r and t. In fact, it is known for $z \le \max_{r,t}[b(r,t)]$ and is truncated at $b(r,t)$ as needed. All cases match at $t=0$, but r_d in case (ii) demonstrates a strong correlation with tide $\langle h_t, r_d \rangle = 0.88$, with minima during low tide and maxima at high tide. Case (iii) is much more variable in Fig. 9(a) and has no strong single frequency component, from the Nyquist period to the inertial period, 0.5–17 h. The significant decline in r_d between 24 and 28 h is due to simultaneous upwelling on the seaward side of the break and downwelling on the shallow side. Internal waves tend to reduce detection range relative to case (ii), by 1–2 km for 200–250 Hz. The standard deviation over the 200–250 Hz

band for case (iii) in Fig. 9(b) also shows an increase over the same period. Otherwise, s_d in both cases (ii) and (iii) indicate no significant periodicity and are usually predicted by the time constant case (i). Results for the detection range within the 300–350 Hz band are illustrated in Fig. 9(c). Case (ii) shows the maxima in r_d occur at mean tide, the minima occur at both high and low tide and the standard deviation in Fig. 9(d) indicates similar variation, but inverted. These variations are closely correlated to the absolute deviation of the tide from the mean, $\langle |h_t|, r_d\rangle = -0.82$ and $\langle |h_t|, s_d\rangle = 0.81$. The variability of r_d for case (iii) is only partially correlated to tide with $\langle h_t, r_d\rangle = 0.43$, although the standard deviation is anticorrelated with a slightly higher magnitude, $\langle h_t, s_d\rangle = -0.59$. The detection range in this band is markedly reduced during upwelling/downwelling events in the intervals of 3–4.5, 12.5–13, 15–16.5, 24.5–25.5, and 28–29 h. It is observed at those times that the TL increases because of the energy being scattered into the bottom. The wavelengths in this band are particularly susceptible to this loss mechanism due to the particular combination of distance (in range) between up and down displacements of the thermocline, magnitude of those displacements and their relation to the break. These combinations occur briefly, but regularly, so that the reductions in r_d are more frequent and of shorter duration than the reduction of r_d in the 200–250 Hz band. The internal waves can reduce detection range up to 5 km relative to the case (ii). In Figs. 9(e) and 9(f), describing the 400–450 Hz band, the result for case (ii) indicates that the r_d and s_d are very closely correlated to tidal height with $\langle h_t, r_d\rangle = 0.82$ and $\langle h_t, s_d\rangle = -0.91$, respectively. For case (iii) in Fig. 9(e), the detection range correlation to tidal variability is lower than in the 300–350 Hz band; the most significant periodic component is centered at 10.2 h, rather than the M2 tidal period of 12.4 h, but the standard deviation shows moderate anti-correlation to tide, $\langle h_t, s_d\rangle = -0.60$. The reductions in r_d for this band occur at virtually the same times as for 300–350 Hz, during the same upwelling/downwelling events. Case (ii) is much more variable in this frequency band, with r_d varying up to 3 km, so that case (iii) only reduces r_d by another 3 km. In summary, higher frequencies show greater decreases in detection range due to tidal variability (up to 2 km over a 30 h time window) and internal waves (an additional 5 km decrease). Also, whereas the tidal height can be used to predict variation in r_d and s_d for a given frequency when $c(z)$ is time independent, the introduction of internal waves indicates that detection range is also dependent on the location and size of the internal waves, often without any discernable periodicity.

2. Azimuthal dependence of the acoustic field

The acoustic results discussed earlier were obtained for a single range-depth slice during periods when there were qualitatively different environments between the source and receiving arrays. Some examples of the azimuthal dependence of the acoustic field on tidally driven stratified flow in a shelf-break region are considered below. A simple exten-
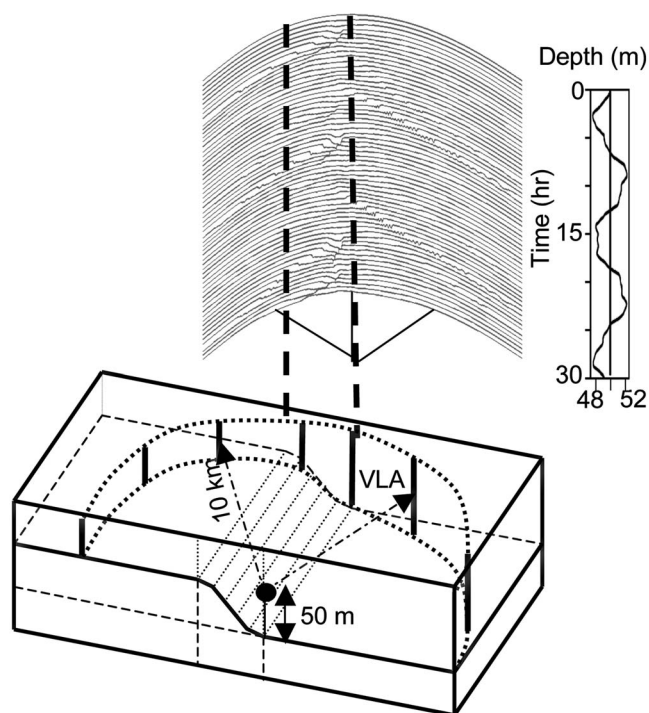
FIG. 10. Source/receiver configuration for 3D simulations of the acoustic field under tidally driven, stratified flow. The source is shown here at the bottom of shelf break and 50 m above the bottom, though results in Figs. 11 and 12 also consider the source positioned in the shelf-break or just at the top of the shelf-break region. The waterfall plot illustrates the azimuth-time distribution of the 1500 m/sec iso-speed contour at a depth of 25 m and using 0.5 h updates. To the right of the waterfall plot is the time signature of the temporal variations in water column depth for the same period is shown. Dotted vertical lines denote the boundaries of the shelf-break region.

sion of the 2D environmental model discussed in Sec. II is made in the along-slope direction (projecting out of the page in Fig. 1. This results in a 3-D shelf-slope region for which both the sound speed field and bathymetry are invariant in the along-slope direction. However, acoustic propagation through this environment exhibits significant azimuthal variability. To illustrate the azimuthal dependence on the acoustic field, a set of 721 vertical arrays is placed on a circular arc within, or adjacent to, the shelf-break region. An example of a source-receiver configuration in this extended environment is depicted in Fig. 10, along with an azimuth-time waterfall display of the 1500 m/s isospeed contour which is located at a depth of approximately 25 m. The arrays span 180° of azimuth and are located at a range of 10 km from bottom moored acoustic sources; these sources are placed in or near the break at midwater depth. The vertical arrays are positioned along the arc in one-third degree azimuthal increments, and extend from the bottom to 10 m below mean water depth. In Fig. 10, the source is located just seaward of the shelf-break and its depth is 50 m from the bottom near midwater depth. A graph of the temporal variation of the water depth on the shelf during a 30 h time period is also displayed in Fig. 10, aligned along the vertical (increasing time) direction of the waterfall plot. The vertical dotted lines delineate the boundaries of the shelf break. This waterfall plot shows a number of internal waves, generated at the shelf break, and radiating out from this region.
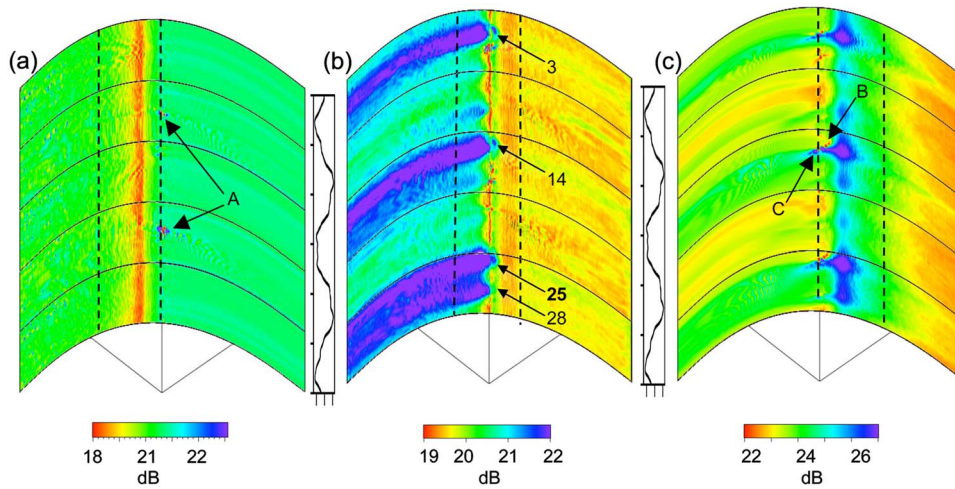
FIG. 11. (Color online) Plots representing depth-averaged transmission loss at 100 Hz, computed over each vertical array located along a 180° circular arc at a range of 10 km from an acoustic source. Results are presented as a function of azimuth and time for three source locations: (a) source at bottom of shelf-break (b) source at midbreak and (c) source at top of shelf break. Dotted vertical lines delineate the shelf-break boundaries. Several features at $t=3$, 14, 25, and 28 h are indicated by the numbered arrows in (b). Similar features appear in (c) though are not numbered.

With Fig. 10 as a reference for the general source-receiver configuration and time evolving environment, the depth-averaged transmission loss over each array is computed at 100 Hz as a function of azimuth and time. The averaged loss is presented in Fig. 11 for three source locations and midwater depths: (a) source at the bottom of the shelf-break, 50 m above bottom, (b) source at midbreak, 38 m above bottom, and (c) source at the top of shelf-break, 25 m above bottom. Note that the scales for each figure are different. The set of parallel curves for fixed times and variable azimuth in Figs. 11(a)–11(c) indicates the time of mean water depth during the 30 h simulation period. The vertical dotted lines again delineate the boundaries of the shelf-break region for the three source positions (the plots are shifted for each source location). Although the bathymetry largely determines the spatial distribution of the transmission loss, some features of the loss can be credited to solibores or other tidal flow induced sound speed variations. The most significant feature of the azimuth-time plot of transmission loss in Fig. 11(a) is the strong focusing located at about midbreak, due to horizontal refraction of acoustic energy along the slope. The levels have weak time dependence in the focusing region, but show little evidence of a tidal signature in this plot. Strong loss is seen at locations labeled A, just seaward of the shelf-break and is associated with significant downwelling of warm water along the propagation path that leads to solibore evolution in deep water. The loss plot for the source at midbreak [Fig. 11(b)] illustrates quite different behavior. In the shallow region (to the left of the dashed lines delineating the shelf break), loss varies approximately with temporal variations in water depth, with higher mean loss at low tide and lower loss at high tide. However, each of these variations in loss is also associated with downwelling over the break [cf. Fig. 2(b)], with the latter generating internal waves in shallow water (cf. Fig. 10). There are four maxima in the transmission loss, occurring at 3, 14, 25, and 28 hours. In particular, during the last low tide from 24 to 30 hr, the loss reaches a maximum twice, at about 25 h (which, notably, precedes tidal minimum) and 28 h. Both times are associated with strong downwelling events at the break. Within the shelf-break region a narrow azimuthal sector exhibits strong, but highly time variable focusing along the isobath

under the source. The focusing is weakest at exactly the same times of higher loss in shallow water, with superimposed shorter period variation of perhaps 3 h. Each reduction can be associated with downwelling. The source is at a fixed distance from the bottom, but the local change in the sound speed profile alters the relative excitation of the modes; it is the likely candidate for the periodic temporal variations in the loss. In particular, downwelling of higher sound speed water to a position below the source depth would lead to lower excitation of the lowest order modes. At the deep end of the shelf break, a broader but relatively time invariant focusing region is present. For a source located above the top of the break at a depth of 25 m [Fig. 11(c)], the propagation exterior to the break region also shows lower intensity during the same time periods as in Fig. 11(b). Indeed, the shadow region at 25 h, which precedes lowest tide but has very strong downwelling, is more intense than at 28 h, which is the low point of that tide. There is also time coincident fading at all azimuths, strongly suggesting reduced source excitation of propagating modes. Additionally, there is focusing (arrow B) and shadowing (arrow C) in the shallow region due to internal wave/bores generated at the shelf break. Along the center of the break, horizontal refraction due to the sloping bottom creates a shadow region, with deeper shadows at the same times as noted earlier.

The relative influence of tidally induced water depth variations and volumetric sound speed perturbations on the azimuthal dependence of the acoustic field is considered next. Fig. 12 shows three sets of paired azimuth-time images which are computed with a generalization of Eq. (5). The modification allows for azimuthal variation of the field, i.e., $\Gamma(\omega, t, t_0, \theta)$ where $0° \leq \theta \leq 180°$. There is one pair of images for each of the source locations. For pair (a,b) the source is positioned at the bottom of the shelf, the source is located at midbreak for pair (c,d) and in pair (e,f) the source is located at the top of the break. The pressure field vectors e and f on each vertical array now have somewhat different meanings than those given in the previous subsection. Here we interpret Eq. (5) on each array as a temporal correlation between pairs of acoustic field vectors defined as follows. For the first image of each pair [Figs. 12(a), 12(c), and
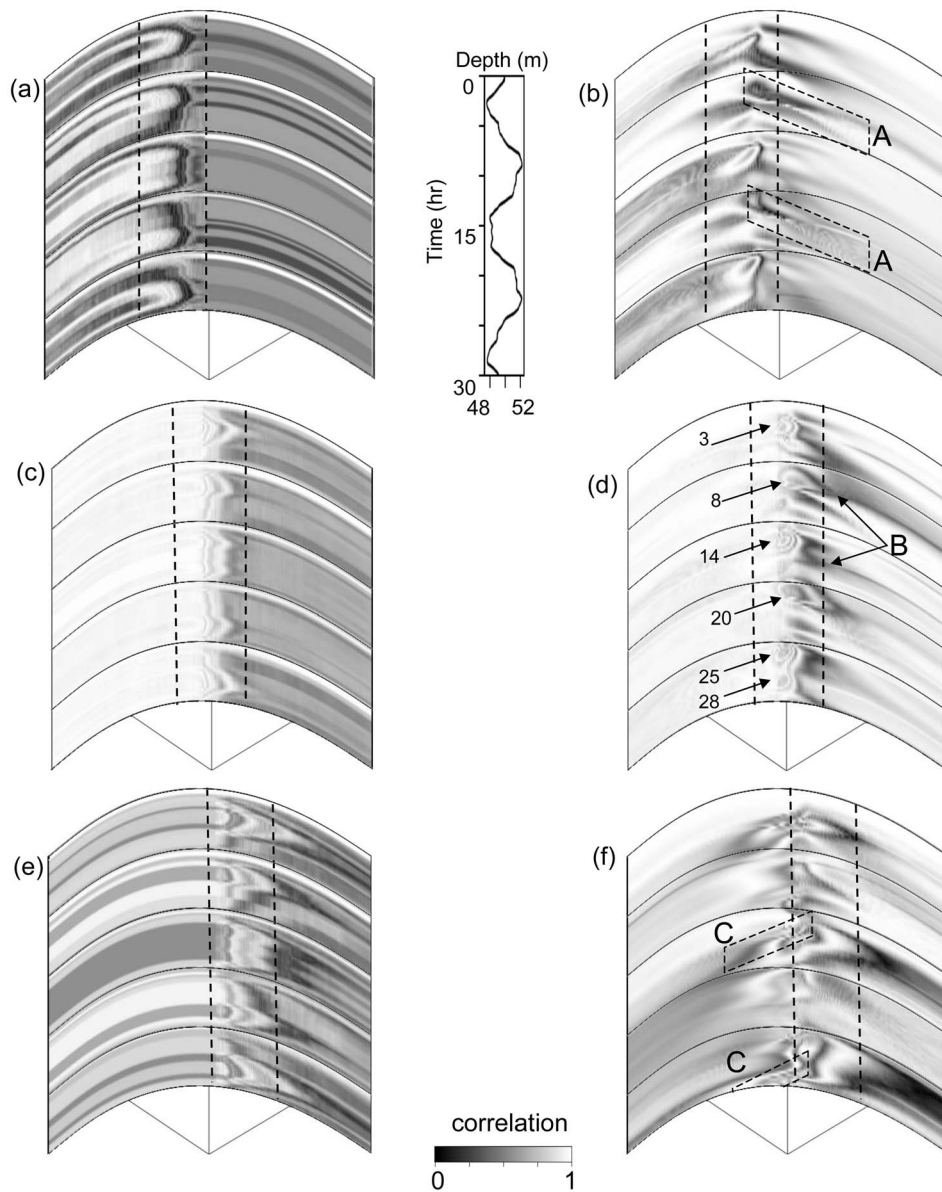
FIG. 12. Azimuth-time correlation plots at 100 Hz, computed over each vertical array located along a 180° circular arc at a source/receiver range of 10 km. There are three sets of paired images, one pair for each source location. The correlated fields displayed in each are discussed in the text. For (a,b) the source is positioned at the bottom of the shelf; in (c,d) the source is located at midbreak; and in (e,f) the source is located at the top of the break. Labeling in (d) is similar to that in Fig. 11(b), and indicates time in hours.

12(e)], the vector f now represents the acoustic field on an array at fixed time $t=0$ h, corresponding to the starting environment in Fig. 2(a). The vector e is now chosen to describe the time evolving acoustic field on an array for the same initial environment though, for $t>0$, the water depth is allowed to change due to tidal variations over the full 30 h period of the computation. Therefore, the correlation between the two pressure vectors in this case depends only on variability in the water column displacement linked to tidal variations and the fields recorrelate perfectly each time the tidal variation passes through the mean height at $t=0$. This does not preclude shorter periods of variation seen on each of these plots, especially Figs. 12(a) and 12(e). No volumetric variations (internal tides and solibore) are included. Note that away from the break, the azimuthal variation is weak or vanishing in Figs. 12(a), 12(c), and 12(e) as the bottom is flat. The correlation in these plots varies significantly, from uniformly near unity in the shelf region of Fig. 12(c) to almost zero at points in the break region of Fig. 12(a). Generally the strongest variation in this set occurs when the propagation

path transverses the entire break as on the shelf in Fig. 12(a) and seaward in Fig. 12(e). The second image of each pair, Figs. 12(b), 12(d), and 12(f), describes the field correlation between the same tidally generated variations of the first image (specified by e above) and f, which now represents the acoustic field on an array for times $t>0$ after it has propagated through the hydrodynamically evolving environments illustrated in Figs. 2 and 10. This correlation then represents a measure of the contribution of internal waves and tides versus what would be expected if the acoustic responses were only due to water depth variations, and will be used to the end of this section. Fig. 12(b) presents a region of rapid variation near the midbreak bearing, the site of the largest advection of the thermocline. The correlation-decorrelation patterns in the break region roughly mimic contours of thermocline displacement for $t<12$ h. As internal waves propagate seaward, (e.g., those generated at $t=9$ or 26 h in Fig. 10), the correlation varies rapidly in the bands within the boxes labeled A in Fig. 12(b). The bands sweep past the arrays near the bottom of the break, corresponding to large

loss [cf. arrows in Fig. 11(a)]; this loss is related to acoustic refraction from internal waves. The bands then fade for bearings away from the break as the direction of acoustic propagation for these bearings more closely aligns with the internal wave's propagation vector. During the low tide periods for $t > 12$ h, the periodic sound speed variation causes correlation to be consistently reduced to about 0.8 on the shelf. In Fig. 12(d), the largest decrease in correlation (arrows labeled B) is associated with downwelling at the deep end of the shelf break. Along the midbreak, ring-like patterns of periodic re-correlation occur at approximately 3, 14, 25, and 28 h, corresponding to the high loss regions seen in Fig. 11(b) where the latter occur during downwelling events. At 8 and 20 h similar elongated patterns occur during downwelling localized over the bottom of the break. In Fig. 12(f), the strongest decorrelation is in deep water at around 14 h and between 25 and 28 h, during high transmission loss periods illustrated in Fig. 11(c). The rapid variation in correlation patterns along the source isobath are less regular than in Fig. 12(d), but remain associated with high loss regions of Fig. 11(c). As in Fig. 12(b), bands of strong, rapid variation of correlation progress with the internal waves' propagation into shallow water, at the boxes labeled C.

IV. SUMMARY AND DISCUSSION

Numerical simulations have been presented for acoustic propagation in a canonical shelf-break ocean environment. They included hydrodynamic specification of the space and time varying sound speed field for the purpose of studying some of the effects of sub-mesoscale oceanography on acoustic field structure. The evolution of the environment was determined in a physically consistent manner with respect to tidal forcing and stratified flows so that internal tides and solibores could be simulated from generation through evolution and propagation. The influence of these oceanographic features on acoustic propagation was illustrated with several 2D and 3D acoustic examples.

The simulations produced acoustic fields propagating within an ocean waveguide where tidally induced water depth variations, internal tides and solibores co-existed and were coupled together. Sample computations describing the frequency and time dependence of signal gain degradation on vertical arrays were presented for two 30 minute time windows. During these periods, the water depth changes induced by the barotropic tide were small (less than .5 m), and this change was reflected in the relatively weak dependence between these variations and SGD. The SGD was dominated by phase rather than amplitude contributions throughout the 200–500 Hz frequency band. There were, however, significant amplitude contributions over several frequency subbands for short range (11 km) propagation during a time interval in which a wave packet crosses the array and the local source environment was approximately time independent. For the short period (30 min) cases, weak degradation was generally observed over the first 2–3 min for the 200–500 Hz band, followed by enhanced degradation whose overall strength exhibited strong frequency and source depth dependence as the environment evolved. Occasionally, frequency selective recorrelation of the signal up to 10 dB was observed. It is interesting to note that both phase dominance of the SGD and recorrelation of the field are consistent with results observed from data and previous modeling efforts[31,33] though no significant effort was made here to model a specific environment. Detection range was estimated for several frequency bands, after performing some spatial averaging of the field intensity. Higher acoustic frequencies were related to decreases in detection range. Smaller decreases in detection range were found when only the tidal variations were present (no internal waves); larger decreases in detection range were obtained when both the tide and internal waves were present, representing the case of fully stratified flow.

Angular (azimuthal) variations in the acoustic field in the vicinity of the shelf break were studied by "extending" the environment in the along-slope direction in conjunction with 3D acoustic computations to propagate the field onto a set of vertical arrays positioned at fixed range and variable azimuth. Depth averaged transmission loss showed several interesting features, including horizontal refraction by the sloping bathymetry and enhanced loss that was time locked to downwelling at the shelf break. Azimuth-time correlation plots were computed for pressure fields on these arrays in order to estimate the relative importance of volume variability and water depth changes, both of which were driven by tidal flow. The sources were placed near or in the break region and the hydrodynamic changes at the break, such as downwelling, dominated the temporal correlation of the acoustic field.

Using the numerical results obtained for the simplified shelf-break environment as a rough guide, one can make some observations concerning sonar performance prediction and phase sensitive signal processors operating in these regions. For example, in shelf-break regions where tidal flow has a significant impact on determining environmental variability, phase-sensitive processors (e.g., matched field or matched mode localization schemes) should adaptively control the frequency band(s) of the replica fields, due to the significant frequency and time dependence of both the degradation and detection range observed here. The acoustic replica vectors would need to be updated in a manner consistent with the rate of signal gain degradation observed in the simulations; for some frequency bands a cyclic update may be appropriate given the observed re-correlation of the field. Because the degradation is a function of M2 tidal phase, the updates should account for this phase-locking behavior. Since the acoustic response in this environment is not shift invariant with respect to space or time, it appears that phase sensitive processors may need to be tuned for each specific environment and source-receiver combination that might be encountered. It should be noted that sonar performance is dependent on other factors not considered here. For example, neither a background (spatially diffuse) internal wave field nor an ambient surface generated noise field were included in this study, though both can negatively impact sonar performance. Other factors that can influence SGD and detection range are acoustic wave scattering from rough bathymetry or bottom inhomogeneity; each can be a contributor to a loss of field coherence. The relative importance

of the surface, volume and bottom contributions is a function of acoustic frequency and is also time and site dependent. While one of these components may dominate under a particular set of environmental conditions, it is often the case that multiple components are found to contribute to the SGD and detection range, and their contributions are not necessarily independent. Additional degrees of freedom due to the introduction of horizontal anisotropy in bathymetry and/or volume variability produces an azimuthal component to the above-mentioned quantities and is reflected in the shift-variant nature of the acoustic field.

Though the ocean model was initiated with a typical summer thermocline profile from the New Jersey shelf, there was no attempt to characterize a specific littoral environment in detail. Although some results are qualitatively consistent with shallow water measurements, considerable care should be exercised in directly relating specific model results to experimental data. This raises the issue of the predictive capability of such models, a subject not directly considered in this paper. Model-data comparisons are impacted by a number of factors. For example, as noted in the Introduction, environmental data are invariably undersampled in space and time; initialization of a numerical (environmental) model typically requires a much finer sampling than can be supplied by the data. The sampling requirements for initialization of submesoscale models are much more difficult to satisfy than those associated with mesoscale models since the latter ignore situations where the ratio of the horizontal to vertical scales in the fluid motion approaches unity. In addition, the finite spatial and temporal scales used in the simulation place limits on acoustic field predictability because they restrict the level of physical description included in the model. For example, this study precluded fine structure and turbulence in the environmental specification because of the relatively coarse grid sampling used in the simulations.

The oceanographic model described here can be fully implemented in three dimensions, though in this paper the flows are restricted to two-dimensional motion. Along-slope variation in the oceanography would introduce additional degrees of freedom (and realism) in the azimuthal dependence of the acoustic field, beyond that illustrated in Sec. III.

Although the above considerations fall outside the scope of this paper, they are points that need to be evaluated in developing a simulation based prediction scheme for sonar performance using vertical or horizontal arrays. Some of the above-mentioned issues can be resolved through modeling enhancements and higher resolution initialization. There exists, however, a deeper issue affecting the general ability of numerical models to make accurate and robust predictions of acoustic field quantities. This issue is related to the question of how one can quantify the lack of knowledge or uncertainty concerning the basic "ingredients" that enter into a numerical simulation. Most numerical modeling studies, including this one, employ inherently deterministic computations of the acoustic field through the solution of a deterministic wave equation. Uncertainty estimates in the boundary conditions, parameters and fields that drive the numerical simulations are not supplied to the algorithm(s). Without such estimates, simulation-based prediction schemes cannot be objectively assessed. Embedding environmental uncertainty directly into the computations is a non-trivial problem.[36,37] In order to develop a robust numerical prediction capability, however, environmental uncertainty should be quantified and included as an integral part of a simulation-based prediction system.

ACKNOWLEDGMENTS

[1] J. R. Apel, M. Badiey, C. S. Chiu, S. Finette, R. Headrick, J. Kemp, J. F. Lynch, A. Newhall, M. H. Orr, B. H. Pasewark, D. Tielburger, A. Turgut, K. von der Heydt, and S. Wolf, "An overview of the 1995 SWARM shallow-water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**, 465–500 (1997).

[2] D. Tielburger, S. Finette, and S. Wolf, "Acoustic propagation through an internal wave field in a shallow water waveguide," J. Acoust. Soc. Am. **101**, 787–808 (1997).

[3] J. Preiseg and T. F. Duda, "Coupled acoustic mode propagation through continental-shelf internal solitary waves," IEEE J. Ocean. Eng. **22**, 256–269 (1997).

[4] B. G. Katsnel'son and S. A. Pereselkov, "Resonance effects in sound scattering by internal wave packets in a shallow sea," Acoust. Phys. **44**, 786–792 (1998).

[5] S. Finette, M. H. Orr, A. Turgut, J. Apel, M. Badiey, C. S. Chiu, R. H. Headrick, J. N. Kemp, J. F. Lynch, A. E. Newhall, K. von der Heydt, B. Pasewark, S. N. Wolf, and D. Tielbuerger, "Acoustic field variability induced by time-evolving internal wave fields," J. Acoust. Soc. Am. **108**, 957–972 (2000).

[6] D. Rubenstein, "Observations of cnoidal internal waves and their effect on acoustic propagation in shallow water," IEEE J. Ocean. Eng. **24**, 346–357 (1999).

[7] M. Badiey, Y. Mu, J. Lynch, J. Apel, and S. Wolf, "Temporal and azimuthal dependence of sound propagation in shallow water with internal waves," IEEE J. Ocean. Eng. **27**, 117–129 (2002).

[8] B. G. Katsnel'son and S. A. Pereselkov, "Low-frequency horizontal acoustic refraction caused by internal wave solitons in a shallow sea," Acoust. Phys. **46**, 684–691 (2000).

[9] M. Badiey, B. G. Katsnelson, J. F. Lynch, S. Pereselkov, and W. L. Siegmann, "Measurement and modeling of three-dimensional sound intensity variations due to shallow-water internal waves," J. Acoust. Soc. Am. **117**, 613–625 (2005).

[10] S. Finette and R. Oba, "Acoustic field variability induced by internal solitary waves on a continental shelf," in *Proceedings of the Institute of Acoustics*, edited by T. G. Leighton, G. J. Heald, H. D. Griffiths, and G. Griffiths (Bath University Press, Bath, UK, 2001), Vol. **23**, pp. 321–326.

[11] R. Oba and S. Finette, "Acoustic propagation through anisotropic internal wave fields: Transmission loss, cross-range coherence, and horizontal refraction," J. Acoust. Soc. Am. **111**, 769–784 (2002).

[12] S. Finette and R. Oba, "Horizontal array beamforming in an azimuthally anisotropic internal wave field," J. Acoust. Soc. Am. **114**, 131–144 (2003).

[13] *Oceanography and Acoustics: Prediction and Propagation Models*, edited by A. R. Robinson and D. Lee (AIP Press, New York, 1994).

[14] R. C. Rhodes, H. E. Hurlburt, A. J. Wallcraft, C. N. Barron, P. J. Martin, E. J. Metzger, J. F. Shriver, D. S. Ko, O. Smedstad, S. L. Cross, and A. Kara, "Navy real-time global modeling systems," Oceanogr. **15**, 29–43 (2002).

[15] K. Lamb, "Numerical experiments of internal wave generation by strong tidal flow across a finite amplitude bank edge," J. Geophys. Res., [Oceans] **99**(1), 843–864 (1994).

[16] K. Winters, "Simulation of non-hydrostatic, density-stratified flow in irregular domains," Int. J. Numer. Methods Fluids **32**, 263–284 (2000).

[17] C. Shen, "Constituent boussinesq equations for waves and currents," J. Phys. Oceanogr. **31**, 850–859 (2001).

[18] C. Y. Shen and T. E. Evans, "A free-surface hydrodynamic model for density-stratified flow in the weakly to strongly nonhydrostatic regime," J. Comput. Phys. **200**, 695–717 (2004).

[19] P. C. Gallacher, S. Piacsek, and D. Dietrich, in *Estuarine and Coastal Modeling*, Proceedings of the 7th International Conference of American Society of Civil Engineers, 2001, pp. 204–214.

[20] S. Finette, R. Oba, P. Gallacher, and M. Schafferkotter, "Acoustic field propagation through a time-evolving buoyant jet," J. Acoust. Soc. Am. **115**, 2550 (2004).

[21] S. Finette, T. Evans, and C. Shen, "Submesoscale modeling of environmental variability in a shelf-slope region and the effect on acoustic fluctuations" in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer Academic, Boston, 2002), pp. 401–408.

[22] S. A. Chin-Bing, A. Warn-Varnas, D. B. King, K. G. Lamb, M. Teixeira, and J. A. Hawkins, "Analysis of coupled oceanographic and acoustic soliton simulations in the Yellow Sea: a search for soliton-induced resonances," Math. Comput. Simul. **62**, 11–20 (2003).

[23] P. K. Kundu, *Fluid Mechanics* (Academic, New York, 1990).

[24] F. J. Millero and A. Poisson, "International one-atmosphere equation of state of seawater," Deep-Sea Res., Part A **28A**, 625–629 (1981).

[25] V. A. Del Grosso, "New equation for the speed of sound in natural waters (with comparisons to other equations)," J. Acoust. Soc. Am. **56**, 625–629 (1981).

[26] M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).

[27] D. Lee and M. H. Schultz, *Numerical Ocean Acoustic Propagation in Three Dimensions* (World Scientific, Singapore, 1995).

[28] A. P. Rosenberg, "A new rough surface parabolic equation program for computing low-frequency acoustic forward scattering from the ocean surface," J. Acoust. Soc. Am. **105**, 144–153 (1999).

[29] M. H. Orr and P. C. Mignerey, "Matched-field processing gain degradation caused by tidal flow over continental shelf bathymetry," J. Acoust. Soc. Am. **111**, 2615–2620 (2002).

[30] R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, J. Apel, M. Badiey, C. S. Chiu, S. Finette, M. Orr, B. Pasewark, A. Turgut, S. Wolf, and D. Tielbuerger, "Acoustic normal mode fluctuation statistics in the 1995 SWARM internal wave scattering experiment," J. Acoust. Soc. Am. **107**, 201–220 (2000).

[31] P. C. Mignerey and M. H. Orr, "Observations of matched-field autocorrelation time in the South China Sea" IEEE J. Ocean. Eng. **29**, 1280–1291 (2004).

[32] C. S. Chiu, "Realistic simulation studies of acoustic signal coherence in the presence of an internal soliton wavepacket," in *The 1998 WHOI/IOS/ONR Internal Solitary Wave Workshop: Contributed Papers*, WHOI Tech Report WHOJ-99-07, 1999.

[33] D. Rouseff, A. Turgut, S. N. Wolf, S. Finette, M. H. Orr, B. H. Pasewark, J. R. Apel, M. Badiey, C. Chiu, R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, and D. Tielbuerger, "Coherence of acoustic modes propagating through shallow water internal waves," J. Acoust. Soc. Am. **111**, 1655–1666 (2002).

[34] D. R. DelBalzo, C. Feuillade, and M. M. Rowe, "Effects of water-depth mismatch on matched-field localization in shallow water," J. Acoust. Soc. Am. **83**, 2180–2185 (1988).

[35] A. Tolstoy, S. Jesus, and O. Rodriguez, "Tidal effects on mfp via the INTIMATE96 test," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer Academic, Boston, 2002), pp. 457–463.

[36] S. Finette, "Embedding uncertainty into ocean acoustic propagation models," J. Acoust. Soc. Am. **117**, 997–1000 (2005).

[37] S. Finette, "A stochastic representation of environmental uncertainty and its coupling to acoustic wave propagation in ocean waveguides," J. Acoust. Soc. Am. **120**, 2567–2579 (2006).

Single-crystal lead magnesium niobate-lead titanate (PMN/PT) as a broadband high power transduction material

Mark B. Moffett[a)]
Anteon Corporation, Middletown, Rhode Island 02842

Harold C. Robinson
Newport Division, Naval Undersea Warfare Center, Newport, Rhode Island 02841

James M. Powers[b)]
EDO Corporation, Salt Lake City, Utah 84115

P. David Baird[b)]
EDO Corporation, Salt Lake City, Utah 84115

Two experimental underwater acoustic projectors, a tonpilz array, and a cylindrical line array, were built with single crystal, lead magnesium niobate/lead titanate, a piezoelectric transduction material possessing a large electromechanical coupling factor ($k_{33}=0.9$). The mechanical quality factor, Q_m, and the effective coupling factor, k_{eff}, determine the frequency band over which high power can be transmitted; k_{eff} cannot be greater than the piezoelectric material value, and so a high material coupling factor is a requisite for broadband operation. Stansfield's bandwidth criteria are used to calculate the optimum Q_m value, $Q_{opt} \sim 1.2 \ (1-k_{eff}^2)^{1/2}/k_{eff}$. The results for the tonpilz projector exhibited $k_{eff}=0.730$, $Q_m=1.17$ (very near optimal), and a fractional bandwidth of 0.93. For the cylindrical transducer array, $k_{eff}=0.867$, $Q_m=0.91$ (larger than the optimum value, 0.7), and the bandwidth was 1.16. Although the measured bandwidths were less than optimal, they were accurately predicted by the theory, despite the highly simplified nature of the Van Dyke equivalent circuit, on which the theory is based. [DOI: 10.1121/1.2717496]

I. INTRODUCTION

Single crystal lead magnesium niobate/lead titanate (PMN/PT), because of its high electromechanical coupling factor (~ 0.9) and high energy density ($\sim 10 \ kJ/m^3$), shows great promise as a transduction material for use in underwater sound projectors.[1] The high coupling factor allows the usable bandwidth (the frequency band over which the transducer impedance is within certain limitations) to be extended dramatically beyond what is achievable with polycrystalline lead zirconate titanate (PZT) transducers.

Following Stansfield's criteria for determining bandwidth,[2] one finds an optimum value, Q_{opt}, for the mechanical quality factor, Q_m, that maximizes bandwidth for a given k_{eff}. If Q_m is greater than Q_{opt}, the band will be limited by excessive variation in admittance magnitude, whereas Q_m less than Q_{opt} results in band limitation by excessive phase variations. In either case, the resulting bandwidth is less than the maximum achievable value for the given k_{eff}, and the transducer designer will try to adjust Q_m to bring it closer to Q_{opt}.

Table I lists the properties of a single-crystal material having a nominal composition of 72% PMN and 28% PT.

The longitudinal-mode coupling factor, k_{33}, for this material is 0.90, which is significantly higher than the k_{33} values of PZT ceramic materials that are in widespread use for underwater acoustic projectors. For example, the k_{33} value[3] for PZT-8 is 0.64.

From the well-known[4–6] relationship between transmitting bandwidth and coupling factor, one should expect to realize an appreciable increase in bandwidth by using single-crystal PMN-PT in underwater acoustic projector applications. However, mere substitution[7] of PMN/PT in place of PZT will not usually result in significant improvement, because the optimum mechanical quality factor, Q_{opt}, for PZT is too high for optimum usage of PMN/PT. Therefore, in designing transducers to exploit the newer material, it is necessary to lower Q_m significantly, and this may require a new design, e.g., with reduced masses, increased compliances, or increased radiation resistance.

In this paper, we present results for two projector arrays: (1) a planar array of tonpilz (longitudinal vibrator) transducers, and (2) a line array of ring transducers, that were designed, not to solve a specific sonar requirement, but to demonstrate the wide transmission bandwidth achievable due to the high coupling factor of single-crystal PMN-PT.

II. THEORY

The simplified equivalent circuit for a piezoelectric projector transducer shown in Fig. 1 is that of Van Dyke.[8] The

---

[a)]Now retired. Author to whom correspondence should be addressed. Electronic mail: marilyn.moffett@verizon.net
[b)]Now with Progeny Systems Corporation, Salt Lake City, UT 84111.

TABLE I. Properties of single-crystal PMN-PT.

Property	Symbol	Value	Units
Mass density	ρ_x	8040	kg/m^3
Bias field		0.44	MV/m
Short-circuit compliance coefficient	s_{33}^E	50	pm^2/N
Free dielectric constant	$\varepsilon_{33}^T/\varepsilon_0$	5000	
Piezoelectric constant	d_{33}	1340	pm/V
Electromechanical coupling factor	k_{33}	0.90	
Short-circuit extensional wave speed	$(\rho_x s_{33}^E)^{-1/2}$	1580	m/s

projector is driven by the voltage, E, which appears across the blocked capacitance, C_b, and the motional branch of the circuit, consisting of a motional capacitance, C_m, a motional inductance, L_m, and the motional resistance, R_m. The transducer resonance frequency is

$$f_0 = [2\pi(L_m C_m)^{1/2}]^{-1} \tag{1}$$

and the conductance at the resonance peak is

$$G_0 = 1/R_m. \tag{2}$$

The mechanical quality factor is

$$Q_m = 2\pi f_0 L_m/R_m = [2\pi f_0 C_m R_m]^{-1}, \tag{3}$$

and the effective coupling factor, k_{eff}, is defined by

$$k_{\mathrm{eff}}^2 = C_m/(C_m + C_b). \tag{4}$$

A tuning inductor, L_t, in parallel with the transducer, is selected to resonate with C_b at f_0. It helps to achieve an efficient transfer of power from the power amplifier to the resistive load, R_m. In what follows we use the normalized frequency variables

$$x = f/f_0, \tag{5}$$

and

$$w = x - 1/x. \tag{6}$$

We also normalize the (tuned) admittance to G_0, obtaining

$$y = g + jb, \tag{7}$$

where the normalized conductance is

$$g = (1 + \eta^2)^{-1} \tag{8}$$

and the normalized susceptance is
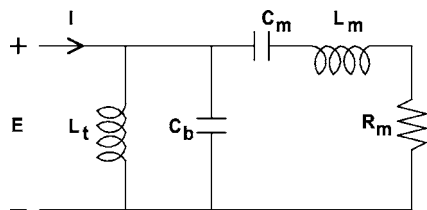
FIG. 1. Van Dyke equivalent circuit showing tuning inductance, L_t, blocked capacitance, C_b, motional capacitance, C_m, motional inductance, L_m, and motional resistance, R_m.
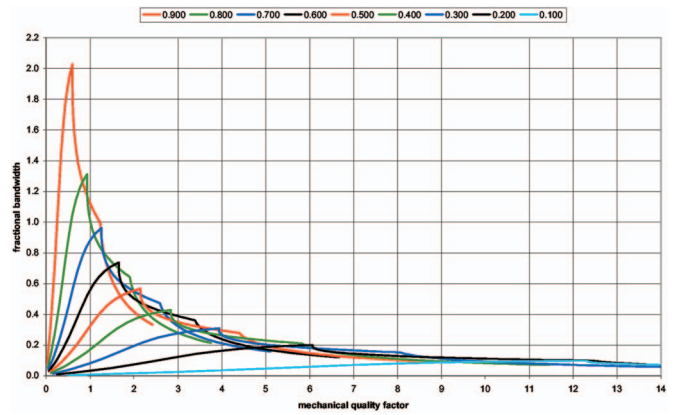
FIG. 2. Bandwidth satisfying Stansfield criteria as a function of mechanical quality factor, Q_m, and effective coupling factor, $k_{\mathrm{eff}}=0.1, 0.2, \ldots, 0.9$.

$$b = [\gamma^{-2} - (1 + \eta^2)^{-1}]\eta, \tag{9}$$

where

$$\eta = wQ_m \tag{10}$$

and

$$\gamma = k_{\mathrm{eff}} Q_m (1 - k_{\mathrm{eff}}^2)^{-1/2}. \tag{11}$$

(In terms of a more familiar parameter, $\gamma^2 = Q_m/Q_Q$).

The tuned transducer forms a bandpass filter between the amplifier and the load. By considering the impedance variation that a properly sized amplifier can tolerate, Stansfield[2] has provided two criteria, for the magnitude and phase of the transducer admittance, to define the transducer bandwidth. The magnitude criterion requires that the variation in magnitude of the tuned admittance be within a factor of 2. The phase criterion restricts the phase variation to $\pm 37°$, i.e., that the phase tangent lie within ± 0.75. (The latter criterion is equivalent to a power factor greater than or equal to 80%.) The useful bandwidth is then determined by whichever criterion is the more stringent. These calculations are given in detail in the appendix. The results are plotted as the multiple curves of Fig. 2, where the bandwidth, BW, is given as a function of the mechanical quality factor, Q_m, and the effective coupling factor, k_{eff}. For a given value of k_{eff}, there is an optimum quality factor

$$Q_{\mathrm{opt}} = 1.245(1 - k_{\mathrm{eff}}^2)^{1/2}/k_{\mathrm{eff}} \sim 1.2(1 - k_{\mathrm{eff}}^2)^{1/2}/k_{\mathrm{eff}}, \tag{12}$$

that results in maximum bandwidth, and that bandwidth is

$$\mathrm{BW}_{\mathrm{max}} = 0.98 k_{\mathrm{eff}}/(1 - k_{\mathrm{eff}}^2)^{1/2} \sim k_{\mathrm{eff}}/(1 - k_{\mathrm{eff}}^2)^{1/2}. \tag{13}$$

The approximate values, 1.2 and unity, in Eqs. (12) and (13), respectively, are commonly used because the more precise values cannot be justified for most transducers, whose behavior deviates from that of the ideal Van Dyke equivalent circuit. As noted by Stansfield, the square root factor in Eqs. (12) and (13) is often ignored, and this is certainly justified if k_{eff} is small, say <0.4. However, use of single-crystal PMN/PT with its larger values of coupling factor requires that we keep the square-root factor in Eqs. (12) and (13).

Figures 3 and 4, respectively, are plots of the optimum mechanical quality factor, Q_{opt}, and the maximum bandwidth $\mathrm{BW}_{\mathrm{max}}$, as functions of the effective coupling factor, k_{eff}. The possibility of large bandwidths from large coupling factors is
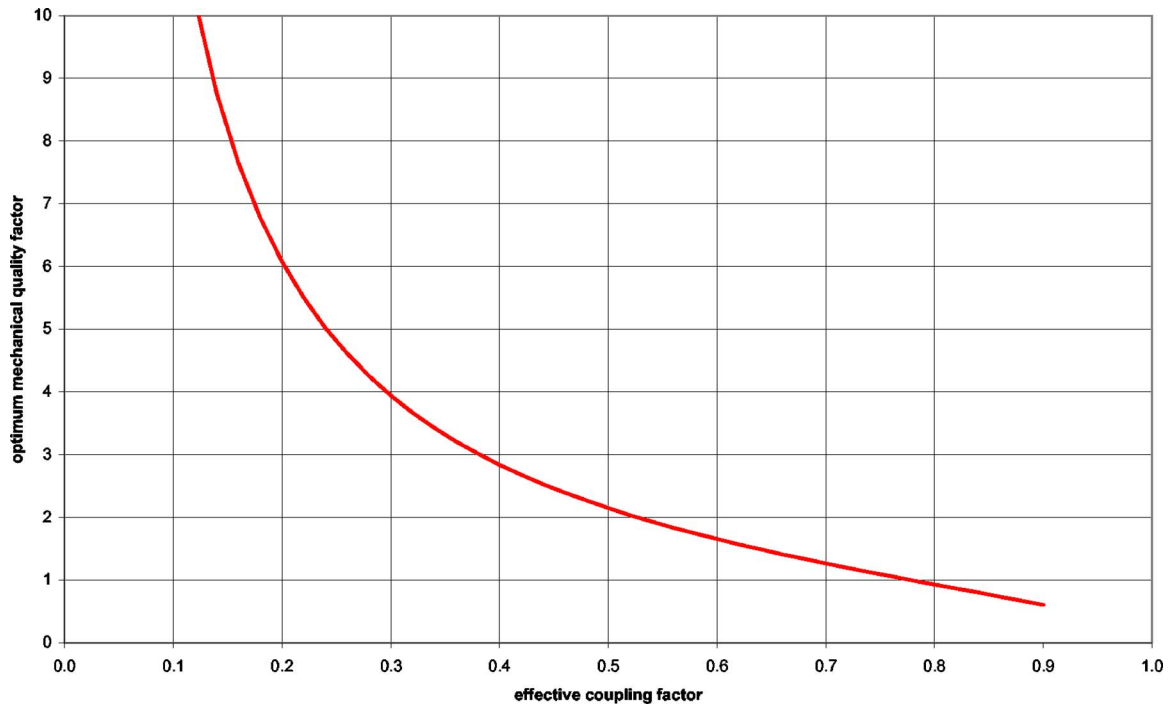
FIG. 3. (Color online) Optimum Q_m for bandwidth, Q_{opt}, as a function of effective coupling factor, k_{eff}.

evident from these plots. For example, if the coupling factor were 0.8 and the mechanical quality factor could be made as low as 0.9, a bandwidth of 1.3 would be achievable. (A fractional bandwidth of 1.3 corresponds to a frequency ratio, $f_+/f_- = 3.4$.)

III. EXPERIMENT

Single-crystal PMN/PT is electrostrictive, with a very small remanent polarization, and so it must be electrically biased. Figure 5 is a simplified schematic diagram of the bias and drive circuitry. The dc-bias supply is shown as a battery of voltage, E_{dc}, and the ac drive is shown as a constant-voltage source of rms amplitude, E. The resistor, R_0, prevents the application of ac to the dc power supply and limits the dc current and power consumed in the dc circuit. The capacitor, C_0, blocks the bias voltage from the amplifier output, but passes ac to the transducer. We require $C_0 \gg 1/(2\pi f_0 R_m)$ and $R_0, \gg R_m$ for these filtering tasks.
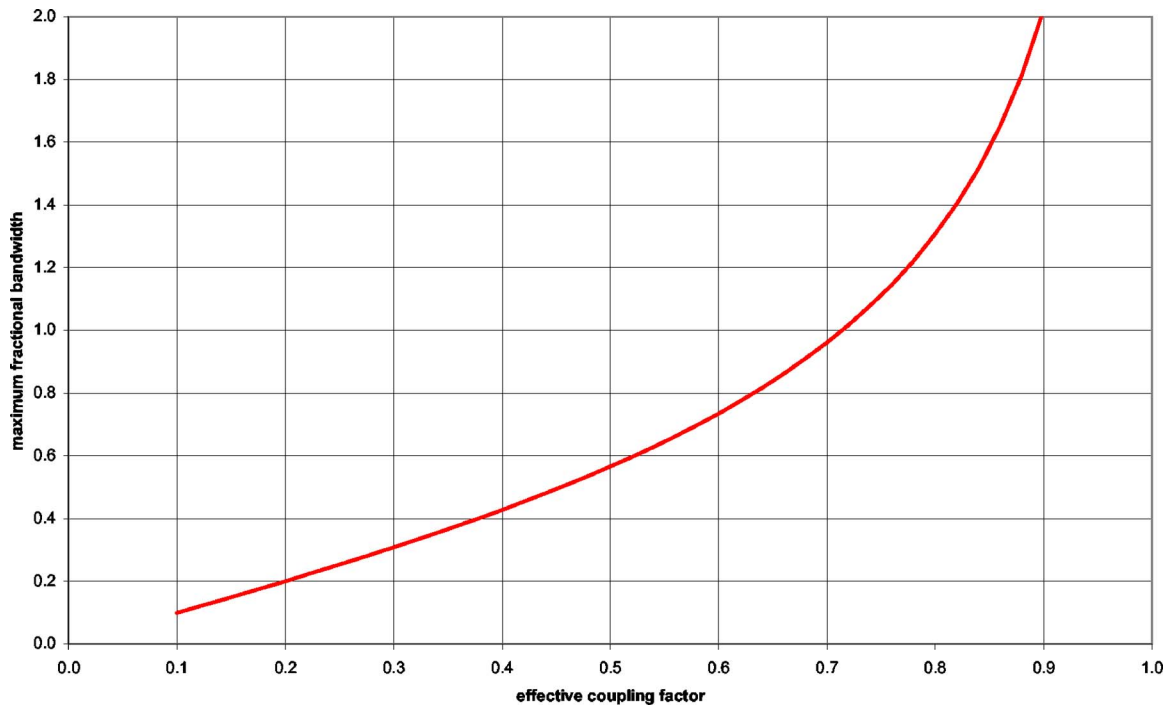
FIG. 4. (Color online) Maximum fractional bandwidth, BW_{max}, as a function of effective coupling factor, k_{eff}.
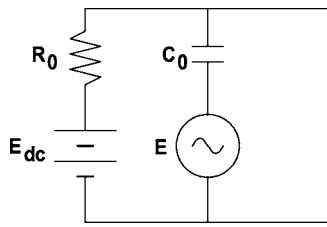
FIG. 5. Circuit for applying bias voltage, E_{dc}, to the transducer.

The 16-element tonpilz array was a planar, 4×4 configuration. Each element (see Fig. 6) was driven axially by four stacks of five single-crystal plates in electrical parallel, with the electric field along the direction of motion. The head mass and tie rod were made of AlBeMet 162, a metal matrix alloy[9] consisting of 38% beryllium and 62% aluminum. Mechanical isolation between the head and tail masses was provided by a Belleville disk spring. Each transducer was supported on the array mounting plate by a syntactic foam isolator. A sheet of polyurethane covering the radiating head masses of the elements served as an acoustic window and prevented water entry. Underwater acoustic measurements were performed in the far field of the tonpilz array in the Naval Undersea Warfare Center's (NUWC's) Acoustic Pressure Tank Facility. The water temperature was 17.6 °C, and the gauge pressure was 4 psig.

The cylindrical array consisted of nine ring-like elements separated axially by thin polyurethane isolators. Each "ring" was actually an air-backed dodecagon of 12 beveled single-crystal plates, with the electric field and the motion along the "circumferential" direction (see Fig. 7). The plates were connected in parallel by means of electrodes attached to the beveled edges between the plates. The rings were each wrapped with glass-fiber-and-epoxy roving to provide compressive prestress along the circumferential direction. After assembly into the array, the elements were inserted in a castor oil filled Tygon tube that served as an acoustic window and seal against water intrusion. Far-field underwater acoustic measurements were performed in the large acoustic tank facility at NUWC at a depth of 8.2 m. The water temperature was 19.3 °C.

FIG. 7. Schematic drawing of cylinder element. Arrows indicate polarization direction in the beveled crystal plates.

IV. RESULTS

Figure 8 shows the normalized conductance (red curves) and susceptance (green curves) as functions of the normalized frequency, f/f_0, for the tonpilz array. The susceptance of the connecting cable was subtracted from the raw data, so that the transducer behavior itself could be more closely examined. The heavier curves are the experimental data, while the lighter curves correspond to the theory for $Q_m = 1.17$ and $k_{eff} = 0.730$, values which were obtained from the conductance and susceptance data shown. (Since the Van Dyke circuit exhibits only one degree of freedom, it does not account for the spurious resonances near $0.7f_0$ and $1.6f_0$.) Substitution of these values into Eq. (11) yields $\gamma = 1.25$, i.e., very near the optimum value of 1.24 for maximum bandwidth.

Figure 9 shows normalized admittance loops for the tonpilz transducer array, tuned to resonance by means of a parallel inductance. The heavier loop is the experimental data,

FIG. 6. Schematic drawing of tonpilz element. Arrows indicate polarization direction in the stacked crystal plates.
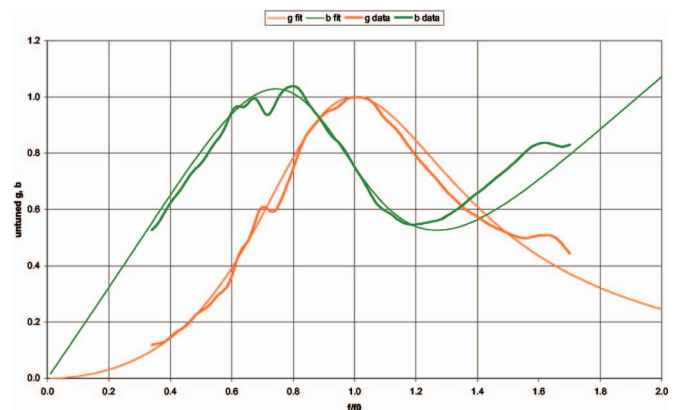
FIG. 8. Normalized, untuned conductance (red) and susceptance (green) of tonpilz transducer. Heavier curves: experimental data. Lighter curves: theory for $Q_m = 1.17$ and $k_{eff} = 0.730$.

Moffett *et al.*: Single-crystal broadband power transduction material
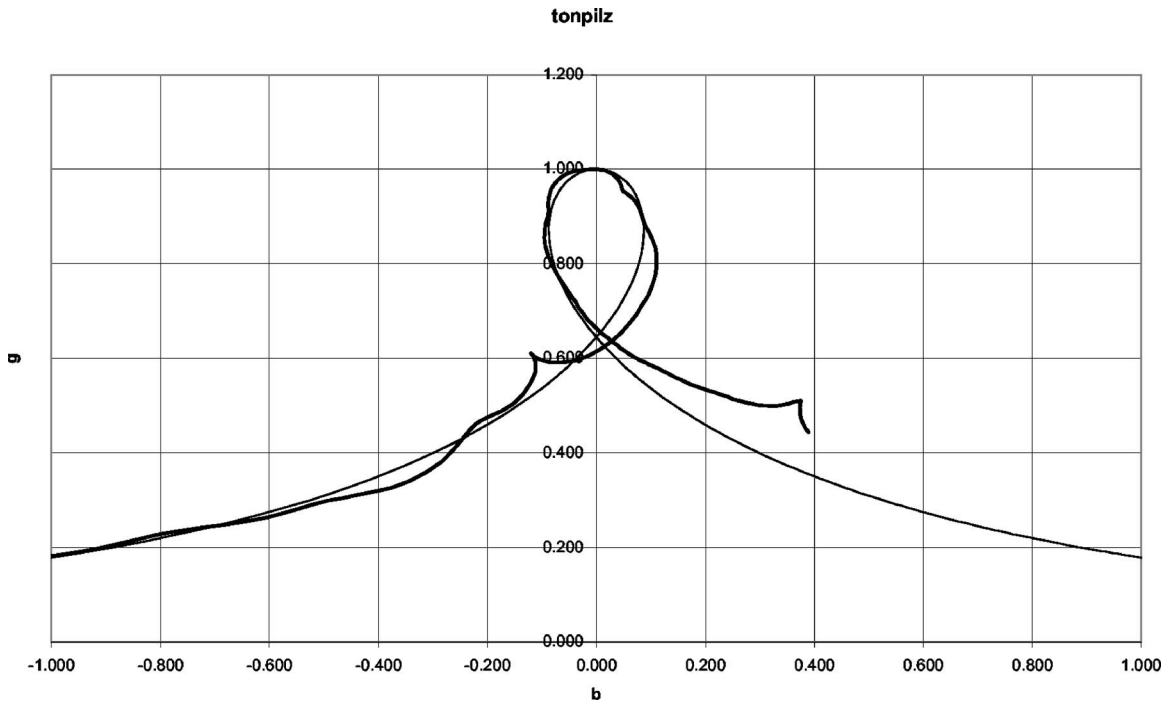
FIG. 9. Normalized, tuned admittance loop for tonpilz transducer array. Heavier curve: experimental data. Lighter curve: theory for $\gamma = 1.25$.

while the lighter one is the theoretical result for $\gamma = 1.25$. The bandwidth can be obtained from the data of Fig. 10, which shows the normalized admittance magnitude (red curve) and the power factor (green curve) as functions of the normalized frequency, f/f_0, for the tonpilz array. Using the Stansfield criteria, we determined the band limits to be 0.635 f_0 and 1.565 f_0, where both limits are based on the magnitude criterion. Therefore the bandwidth of the tonpilz transducer array was the difference between these two limits, namely, BW=0.93, i.e., about 12% less than $BW_{max} = 1.06$ calculated from Eq. (13).

The data for the cylindrical transducer array are given in Figs. 11–13. Figure 11 shows the untuned, normalized conductance (red) and susceptance (green) with the cable removed analytically. Analysis of the heavy curves of Fig. 11 yields $Q_m = 0.91$ and $k_{eff} = 0.867$. Substitution of these values into Eq. (11) yields $\gamma = 1.58$. Because the Van Dyke circuit

does not take the frequency dependence of the radiation load into account, the theoretical fit deviates from the experimental results below f_0. Figure 12 shows experimental and theoretical parallel-tuned admittance loops, where the theoretical loop is for $\gamma = 1.58$. Figure 13 shows the tuned admittance magnitude (red) and power factor (green). Application of the Stansfield criteria to the experimental (heavier) curves of Fig. 13 yields band limits of 0.678 f_0 and 1.836 f_0 for a fractional bandwidth, BW, of 1.16.

The bandwidth results are presented in Fig. 14, where the red curve represents the theory for the cylindrical transducer (i.e., $k_{eff} = 0.867$), and the green curve is that for the tonpilz ($k_{eff} = 0.730$). The two transducer data points, shown as the red and green square symbols, are seen to lie very close to the corresponding curves, with the tonpilz data point (green) actually on the curve.
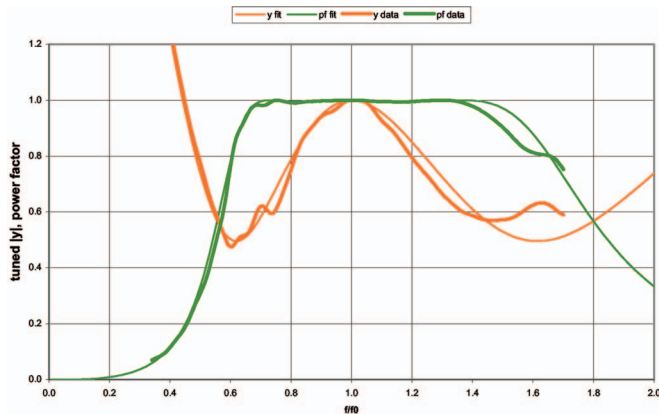
FIG. 10. Tuned admittance magnitude (red) and tuned power factor (green) of tonpilz transducer. Heavier curves: experimental data. Lighter curves: theory for $Q_m = 1.17$ and $k_{eff} = 0.730$.
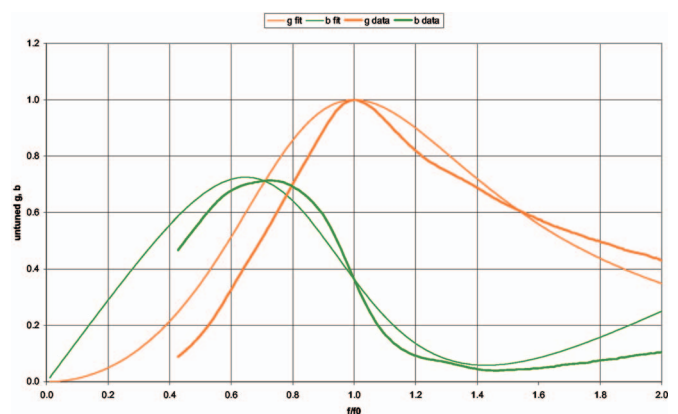
FIG. 11. Normalized, untuned conductance (red) and susceptance (green) of cylindrical transducer. Heavier curves: experimental data. Lighter curves: theory for $Q_m = 0.91$ and $k_{eff} = 0.867$.
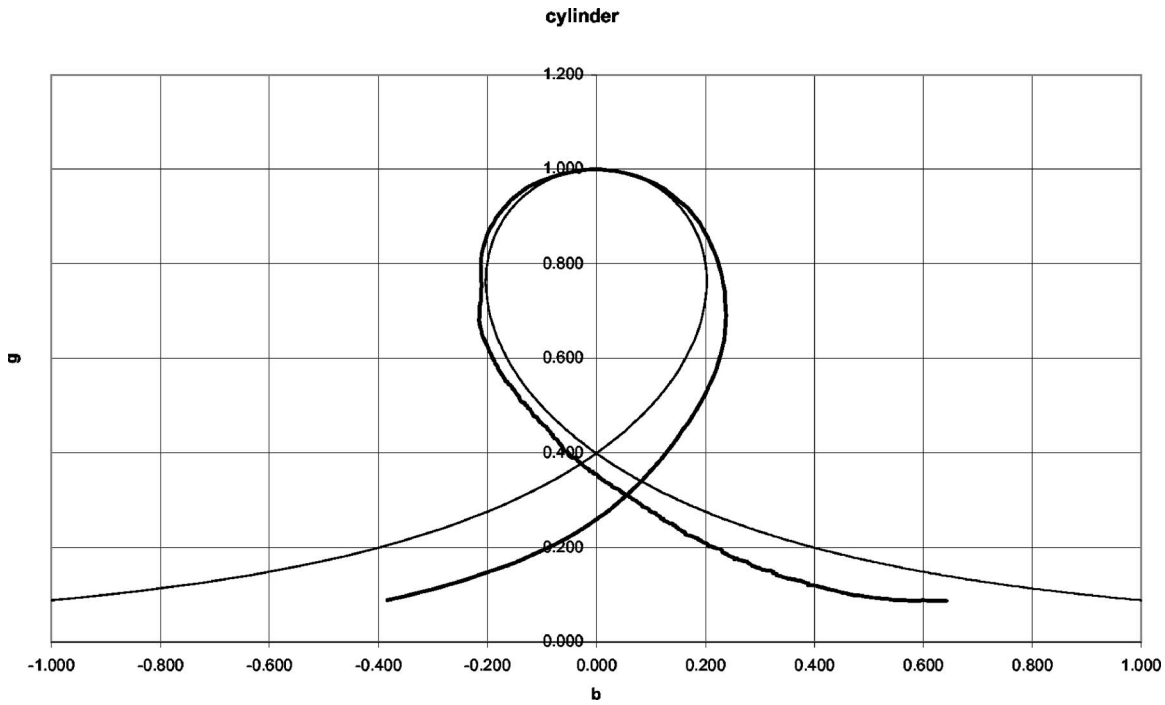
FIG. 12. Normalized, tuned admittance loop for cylinder transducer array. Heavier curve: experimental data. Lighter curve: theory for $\gamma = 1.58$.

V. CONCLUSIONS

It can be concluded from the results shown in Fig. 14 that the theory presented here provides an accurate prediction of the bandwidth to be expected for the two single-crystal transducers, despite the highly simplified nature of the Van Dyke equivalent circuit, on which the theory is based. The theory makes no attempt to account for the spurious resonances exhibited by the tonpilz array or the frequency dependence of the radiation impedance that is believed to cause the skewed shape of the cylinder array response. Still, as a first approach to achieving the optimum bandwidth, the simplified theory seems to be a reliable guide. In order to achieve the maximum bandwidth from a given effective coupling factor (for which reasonable estimates can usually be made during the design process) one needs to strive for a mechanical quality factor satisfying Eq. (12). In the present instance, we succeeded more with the tonpilz array than with the cy-

lindrical array. In the case of the latter, $Q_{\text{opt}} \sim 0.7$, the optimum value, was just too low to meet within the size constraint for the device. The use of single-crystal PMN/PT, with its large material $k_{33} = 0.9$, compared to 0.7 for PZT ceramic materials, allows one to achieve greater bandwidth than with PZT, if one can reduce the Q_m accordingly. This can be accomplished by a combination of decreasing the motional mass, increasing the motional compliance, and increasing the motional resistance, R_m, of the projector. (One could, of course, increase R_m by increasing the mechanical damping, but that would have the undesirable effect of a lowered efficiency, and so the preferred way to increase R_m is to increase the radiation resistance, e.g., by increasing the radiating area.) Figure 4 shows the dramatic increases in bandwidth than can be achieved. For example, if a typical PZT-8 transducer coupling factor is 0.5, with a maximum
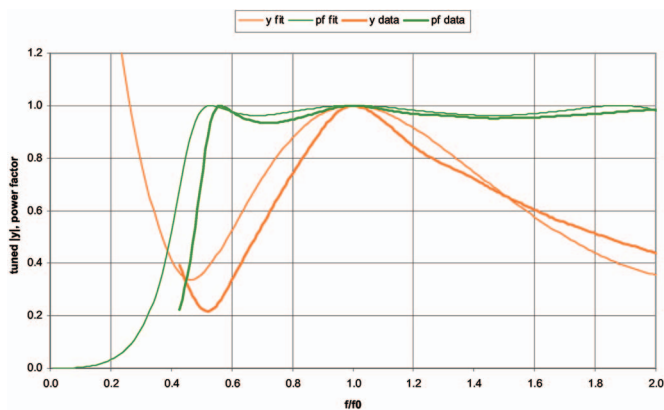
FIG. 13. Tuned admittance magnitude (red) and tuned power factor (green) of cylinder transducer array. Heavier curves: experimental data. Lighter curves: theory for $Q_m = 0.91$ and $k_{\text{eff}} = 0.867$.
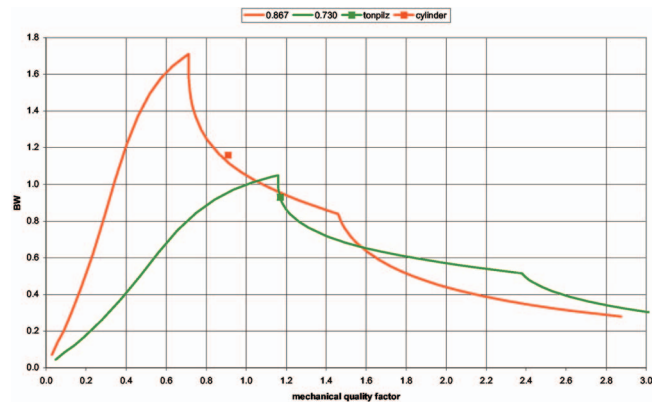
FIG. 14. Fractional bandwidth as a function of mechanical quality factor for $k_{\text{eff}} = 0.867$ (red) and 0.730 (green). Theory shown as solid curves. Experimental data shown as square symbols: cylinder array (red) and tonpilz array (green).

Moffett *et al.*: Single-crystal broadband power transduction material

fractional bandwidth ~ 0.6, and one achieves a coupling factor of 0.75 with single-crystal PMN/PT, the fractional bandwidth could be almost doubled to ~ 1.1.

GLOSSARY OF SYMBOLS

b = normalized tuned susceptance, B/G_0
B = tuned susceptance [S]
BW = fractional bandwidth, normalized to f_0
$\mathrm{BW}_{1/2}$ = half-power fractional bandwidth, $x_+ - x_-$.
$\mathrm{BW}_{\max}$ = maximum fractional bandwidth, see Eq. (13)
C_0 = bias-blocking capacitance, see Fig. 5 [F]
C_b = blocked capacitance, see Fig. 1 [F]
C_m = motional capacitance, see Fig. 1 [F]
d_{33} = piezoelectric constant [m/V]
E = rms drive voltage [V]
E_{dc} = dc bias voltage [V]
f = frequency [Hz]
f_0 = short-circuit resonance frequency, see Eq. (1) [Hz]
f_- = frequency at lower band limit [Hz]
f_+ = frequency at upper band limit [Hz]
g = normalized tuned conductance, G/G_0
G = tuned conductance [S]
G_0 = conductance at the resonance peak, see Eq. (2) [S]
I = rms drive current [A]
j = $(-1)^{1/2}$
k_{33} = electromechanical coupling factor
k_{eff} = effective coupling factor, see Eq. (4)
L_m = motional inductance, see Fig. 1 [H]
L_t = tuning inductance [H]
Q_m = mechanical quality factor, see Eq. (3)
Q_{opt} = optimum Q_m, see Eq. (12)
R_0 = bias-current-limiting resistor, see Fig. 5 [Ω]
R_m = motional resistance, see Fig. 1 [Ω]
s_{33}^E = short-circuit compliance coefficient [Pa^{-1}]
w = frequency variable, $x - 1/x$
w_- = value of w at lower band limit
w_+ = value of w at upper band limit
x = normalized frequency, f/f_0
x_- = normalized frequency at lower band limit, f_-/f_0
x_+ = normalized frequency at upper band limit, f_+/f_0
y = normalized tuned admittance, $g + jb$
γ = coupling parameter, $k_{\mathrm{eff}} Q_m (1 - k_{\mathrm{eff}}^2)^{-1/2}$
$\varepsilon_{33}^T/\varepsilon_0$ = free dielectric constant
η = frequency variable, $w Q_m$
ρ_x = mass density [kg/m^2]

APPENDIX: BANDWIDTH CALCULATIONS

A commonly used definition of bandwidth[10] would be the frequency separation of the half-power points for constant-voltage drive conditions, i.e., the half-conductance points. From Eq. (8), we see that these points occur at $\eta = \pm 1$, i.e.,

$$w_+ = +1/Q_m \tag{A1}$$

and

$$w_- = -1/Q_m. \tag{A2}$$

The corresponding half-power (normalized) frequencies are

$$x_+ = (w_+/2) + [1 + (w_+/2)^2]^{1/2} \tag{A3}$$

and

$$x_- = (w_-/2) + [1 + (w_-/2)^2]^{1/2}. \tag{A4}$$

The half-power bandwidth, expressed as a fraction of the resonance frequency, is then

$$\begin{aligned} \mathrm{BW}_{1/2} = x_+ - x_- &= (w_+/2) + [1 + (w_+/2)^2]^{1/2} \\ &\quad - (w_-/2) - [1 + (w_-/2)^2]^{1/2}, \end{aligned} \tag{A5}$$

but because $w_- = -w_+$, we have

$$\mathrm{BW}_{1/2} = w_+ = 1/Q_m. \tag{A6}$$

Because this definition of bandwidth applies to constant-voltage drive conditions, it is not realistic for practical conditions, such as when a high-power transducer is driven near resonance by an amplifier that is not greatly oversized for its job. Equation (A6) suggests that arbitrarily large bandwidths could be achieved by reducing Q_m. However, as noted by Woollett,[4] the Mason[5] criterion states that the bandwidth over which power can be efficiently delivered to the projector will always be less than or equal to $k_{\mathrm{eff}}/(1 - k_{\mathrm{eff}}^2)^{1/2}$, and Eq. (A6) only satisfies this criterion as long as $Q_m \geq (1 - k_{\mathrm{eff}}^2)^{1/2}/k_{\mathrm{eff}}$. To define the bandwidth for lower values of Q_m and to determine the optimum Q_m that will allow us to achieve the maximum bandwidth, we use the Stansfield criteria defining the locus of the usable area in the admittance plane of the tuned transducer.

Because of the symmetry of g and b in the variable, w, the solutions to the Stansfield criteria always result in a lower band limit, w_-, that is the negative of the upper band limit, i.e., $w_- = -w_+$. Then, as in the example above, the bandwidth becomes simply equal to w_+ itself. The band edges defined by the phase criterion are solutions of
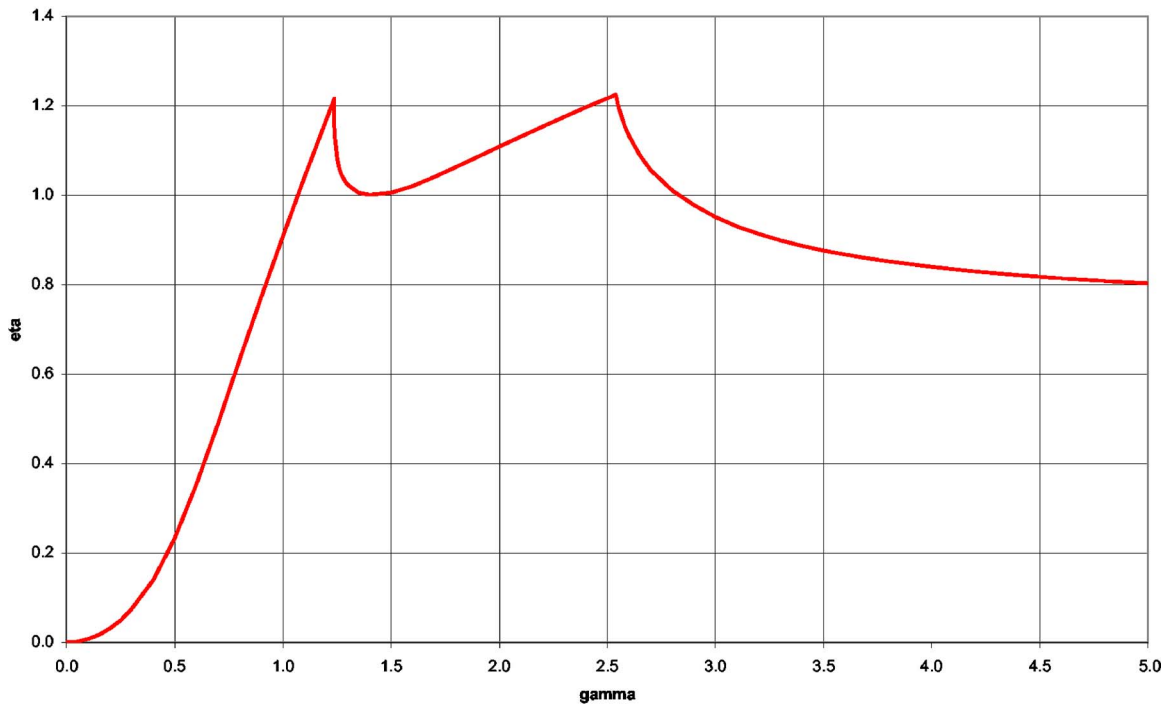
FIG. 15. (Color online) Bandwidth variable, $\eta = wQ_m$, satisfying the Stansfield criteria, plotted as a function of $\gamma = k_{\mathrm{eff}}Q_m(1-k_{\mathrm{eff}}^2)^{-1/2}$.

$$b/g = \pm 0.75. \tag{A7}$$

Use of Eqs. (8) and (9) yields

$$[\gamma^{-2}(1+\eta^2) - 1]\eta = \pm 0.75. \tag{A8}$$

The band edges according to the magnitude criterion are solutions of

$$g^2 + b^2 = 0.25, \tag{A9}$$

and use of Eqs. (8) and (9) yields

$$[\gamma^{-2}(1+\eta^2) - 1]\eta = \pm[0.25(1+\eta^2)^2 - 1]^{1/2}. \tag{A10}$$

Figure 15 is a plot of the solution, expressed as $\eta = Q_m w_+$, to the Stansfield criteria in terms of the parameter, $\gamma = k_{\mathrm{eff}}Q_m/(1-k_{\mathrm{eff}}^2)^{1/2}$. The phase criterion gives rise to the left and right ends of the curve, while the magnitude criterion yields the middle portion. In other words, the magnitude criterion is the more stringent of the two criteria for $1.24 < \gamma < 2.54$, whereas the phase criterion is more stringent at higher and lower values of γ. Finally, to obtain the
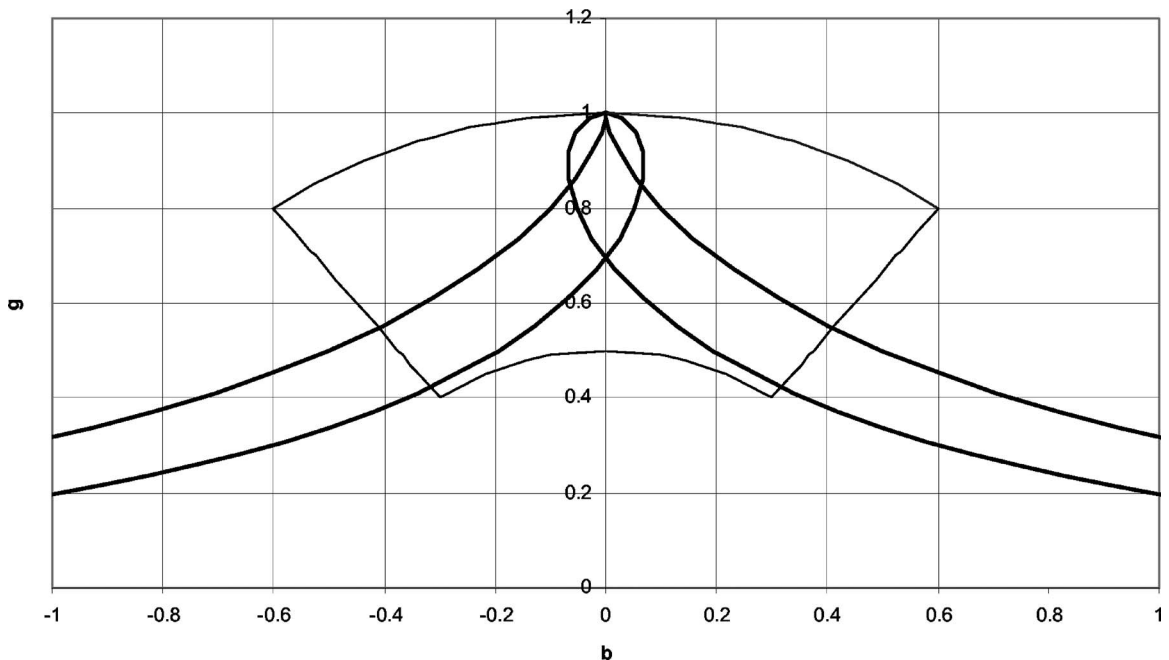
FIG. 16. Normalized tuned admittance loops for $\gamma = 1.0$ and 1.2. Note the cusp for $\gamma = 1.0$.

Moffett *et al.*: Single-crystal broadband power transduction material

multiple bandwidth plots of Fig. 2, we simply multiply the ordinate, η, and abscissa, γ, of Fig. 15 by the scale factors, $1/Q_m$ and $(1-k_{\text{eff}}^2)^{-1/2}/k_{\text{eff}}$, respectively.

Figure 15 shows two peaks at the intersections of the phase and magnitude criteria. The locations of these peaks can be obtained by equating the right-hand sides of Eqs. (A8) and (A10). When this is done, we obtain $\eta=(1.50)^{1/2}=1.225$, i.e., the same value for the height of the two peaks at $\gamma=1.245$ and 2.540. The bandwidths represented by the two peaks are:

$$\text{BW} = \eta/Q_m = \eta k_{\text{eff}}/[\gamma(1-k_{\text{eff}}^2)^{1/2}]$$
$$= (1.225/1.245)k_{\text{eff}}/(1-k_{\text{eff}}^2)^{1/2} = 0.98k_{\text{eff}}/(1-k_{\text{eff}}^2)^{1/2}$$

and

$$(1.225/2.540)k_{\text{eff}}/(1-k_{\text{eff}}^2)^{1/2} = 0.48k_{\text{eff}}/(1-k_{\text{eff}}^2)^{1/2},$$

respectively. Because the first peak, at $\eta=1.225$ and $\gamma=1.245$, represents the greater bandwidth, it constitutes the optimum situation expressed by Eqs. (12) and (13).

It should be noted that the optimum conditions expressed by Eqs. (12) and (13) are a result of adopting the Stansfield criteria. Acceptance of a different region in the complex admittance plane for the bandwidth criteria would change the optimum conditions. For example, there is an optimization approach[4,5,10] that requires equality of the mechanical and electrical quality factors, so that $Q_{\text{opt}}=(1-k_{\text{eff}}^2)^{1/2}/k_{\text{eff}}$, i.e., $\gamma=1.0$ instead of 1.2. The resulting bandwidth, however, turns out to be the same, namely, BW_{max} $\sim k_{\text{eff}}/(1-k_{\text{eff}}^2)^{1/2}$. It is interesting to compare the shapes of the admittance loops for $\gamma=1.0$ and $\gamma=1.2$. This is done in Fig. 16, which also shows the area defined by the Stansfield criteria. The loop for $\gamma=1.0$ is actually a cusp, and for a given coupling factor, will require a 20% lower value for Q_{opt} than will the Stansfield criteria, making the transducer designer's goal a more difficult one to meet.

[1]S. E. Park and T. R. Shrout, "Characteristics of relaxor-based piezolelectric single crystals for ultrasonic transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1140–1147 (1997).

[2]D. Stansfield, *Underwater Electroacoustic Transducers* (Bath University Press, Bath, and Institute of Acoustics, St. Albans, UK, 1990), Chap. 5.

[3]D. Berlincourt, "Piezoelectric crystals and ceramics," in *Ultrasonic Transducer Materials*, edited by O. E. Mattiat (Plenum, New York, 1971), Chap. 2, p. 102.

[4]R. S. Woollett, *The Flexural Bar Transducer*, edited by L. C. Maples (Scientific and Engineering Studies, Naval Underwater Systems Center, Newport, RI, 1986), p. 20.

[5]W. P. Mason, *Electromechanical Transducers and Wave Filters*, 2nd ed. (D. Van Nostrand, Princeton, NJ, 1948), Secs. 6.3, 6.31.

[6]M. B. Moffett and W. J. Marshall, Jr., "The importance of coupling factor for underwater acoustic projectors," NUWC-NPT Technical Document No. 10,691 (Naval Undersea Warfare Center, Newport, RI, 7 June 1994). Also available as NTIS document No. AD-A284309.

[7]R. J. Meyer, Jr., T. C. Montgomery, and W. J. Hughes, "Tonpilz transducers designed using single-crystal piezoelectrics," *Proc. Oceans '02 MTS/IEEE*, Biloxi, MI, 29–31 October 2002, Vol. 4, pp. 2328–2333.

[8]K. S. Van Dyke, "The piezoelectric resonator and its equivalent network," Proc. IRE **16**, 742-764 (1928).

[9]*Beryllium Metal Matrix Composite Avionics Materials* (Brush Wellman, Elmore, OH, 1999).

[10]C. H. Sherman and J. L. Butler, *Transducers and Arrays for Underwater Sound* (Springer, New York, 2007), Sec. 2.8.3.

Acoustic modes propagating along the free surface of granular media

V. Aleshin
Laboratoire d'Acoustique de l'Université du Maine UMR-CNRS 6613, Université du Maine,
Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

V. Gusev[a)]
Laboratoire de Physique de l'Etat Condensé UMR-CNRS 6087, Université du Maine,
Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

V. Tournat
Laboratoire d'Acoustique de l'Université du Maine UMR-CNRS 6613, Université du Maine,
Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

In unconsolidated granular materials under gravity there exist acoustical waves propagating along the surface with anomalously low sound velocity. The presented theory describes these guided surface acoustic modes (GSAM) confined between the surface of the granular materials and in-depth layers with increasing rigidity. The analysis is based on the obtained original analytical solution of the Helmholtz equation that has never been used both in classical and quantum mechanics. This solution is valid for a particular rigidity profile, whereas the general case of grains with or without adhesion has been analyzed numerically. In contrast to the Rayleigh wave polarized in the sagittal (vertical) plane, which is the unique localized mode in a homogeneous solid, an infinite number of modes with sagittal polarization as well as an infinite number of shear horizontal modes have been found. The difference in physical mechanisms of localization is discussed, and the transformation of the GSAMs into the Rayleigh wave at the increasing adhesion is demonstrated: The first sagittal mode transforms into the Rayleigh one, while the others delocalize. The theory explains the experimentally observed magnitude of velocity for the acoustic waves in sand elliptically polarized in the sagittal plane. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2714923]

I. INTRODUCTION

The surface acoustical waves are important both for understanding of the fundamental physical properties of the material surfaces (like heat capacity and thermal conductivity[1]) and for various applications including nondestructive testing and signal processing.[2] Surface phenomena in granular mechanics, such as avalanches, dunes (barchans), are widely addressed in literature.[3–5] On the other hand, there is a vast amount of publications on bulk acoustic modes in granular materials, describing increasingly complicated situations accounting for medium microinhomogeneity, the existence of force chains, jamming, fragility,[6–8] and the intrinsic nonlinearity caused by intergranular contacts.[6,9] At the same time, the acoustic surface and waveguide modes, even in the simplest limit of continuous linear elasticity, are nearly unexplored.[10–12]

For the unconsolidated granular materials under gravity, in the vicinity of their mechanically free horizontal surface, the velocities of longitudinal (L) and shear (S) bulk waves are commonly approximated as a power law dependence on pressure p: $c_{L,S} = \gamma_{L,S} p^{\alpha}$, and, because of the relation $p = \rho g y$, on the depth coordinate y.[13,14] Here ρ and g denote the nearly

constant material density and the gravity acceleration, respectively. The parameters $\gamma_{L,S}$ depend, in particular, on the elastic properties of the individual grains. The simplest estimate for the power α can be obtained using the Hertz theory: $\alpha = 1/6$. At the same time, there are multiple experiments documenting the same dependence $c_{L,S} = \gamma_{L,S}(\rho g y)^{\alpha}$, but with the powers different from $1/6$, the most frequent observation being $\alpha = 1/4$. The experimental results and some plausible theoretical explanations along with original theoretical ideas for different power-law dependencies of sound velocities on pressure can be found in frequently cited papers[13,15] and recent publications.[6,8,16] In this study we do not address the physical origins of power-law dependencies, but concentrate on their consequences for the existence of the acoustic waves localized near the surface.

For homogeneous solids, the mechanism of wave localization is entirely based on the stress-free boundary conditions. For granular materials under gravity, the vertical stratification of the elastic properties suggests another mechanism that consists in the upward bending of the acoustic rays ("mirage" effect[17]) in the near surface layers of the material (see Fig. 1). On the other hand, waves are reflected from the surface as well, as it is predicted by the wave theory in inhomogeneous nonabsorbing media[10,11,18] having the power-law velocity profiles with $\alpha < 1$, even at the normal wave incidence. At the same time, for $\alpha \geqslant 1$ (which is not the

[a)]Electronic mail: Vitali.Goussev@univ-lemans.fr
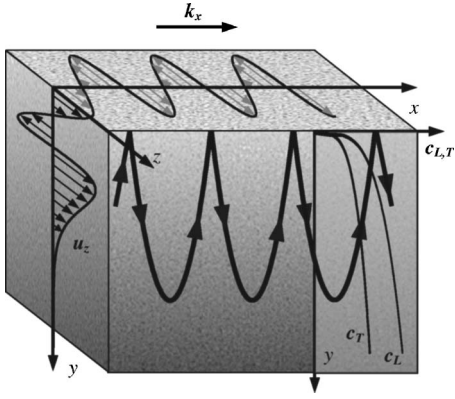
FIG. 1. Qualitative presentation of an unconsolidated granular material with gravity-caused depth (y) variations of elastic properties, shown with a gradient filling and as a power-law depth dependence of sound velocities c_L and c_T at the right. The thin arrows indicate the particle displacements for the SH modes both in the volume and at the surface. The path of a ray in GA approximation (thick lines with arrows) is also shown.

case we analyze here) the upward bending of the rays in the geometrical acoustic approximation[18,19] still occurs, but the reflected wave never reaches the surface since its propagation time becomes infinite (acoustic "black hole"[20,21]).

Neglecting the surface radiation of sound into the air, we conclude that acoustic rays could propagate in the waveguide mode, since they are confined between the free surface on the one hand, and the increasingly rigid material (the caustics[2,10,11,18,19]) on the other one. The experimental observations of such elliptically polarized in the sagittal plane guided surface acoustic modes (GSAMs or GSA modes) propagating along the surface of dry sand have been reported.[22–24] However the measured very low propagation speed of these waves has remained unexplained,[24] because of the absence of an adequate theory. Note that the existing theory[10,11] is valid for the waves localized at the interface of water-saturated sand with water and not for the waves localized at mechanically free surface of a dry sand.

Here we present the results of the theoretical evaluation of possible linear, i.e., of infinitely small amplitude, acoustic modes localized near mechanically free surface of the unconsolidated granular material, which is treated as being perfectly continuous and elastic. The dispersion relations for both shear horizontal (SH) waves and polarized in the sagittal (vertical) plane waves (P+SV), the latter being composed of continuously interacting longitudinal (P) and shear vertical (SV) components, are predicted and the magnitude of the wave velocity observed in the experiment[24] is explained. The developed theory provides an opportunity to predict the dependence of the localized waves' behavior on the adhesion between grains or on the external loading additional to gravity.

In Sec. II, we write out the Helmholtz equations and the boundary conditions for the SH and P+SV modes. Section III is devoted to the analysis of the SH waves through available analytical solutions. A semianalytical method using the Laguerre polynomials and characteristic determinants is applied to both SH and P+SV modes in Sec. IV. In Sec. V we report the results of the direct numerical solution of the eigenproblem in the general case and show the transforma-

tion of the GSAMs into the Rayleigh mode with increasing grain adhesion. The brief report on the most important results of our research has been published elsewhere.[25]

It is worth mentioning that the waves discussed in the following are due to solid-like behavior of the unconsolidated granular material loaded by its own weight, and they are very different from the experimentally observed surface patterns and waves, which are due to fluid-like behavior of the granular material subjected to vibrations.[3,26]

II. THE HELMHOLTZ EQUATION FOR THE LOCALIZED SH AND P+SV MODES

The mechanical displacement field for localized modes must satisfy the form $(u_x, u_y, u_z)\exp[i(\omega t - kx)]$, where x is the coordinate along the propagation direction, ω is the frequency, k is the wave number, u_z is the SH wave component (see Fig. 1), and u_x, u_y are the components of the P+SV wave polarized in the sagittal plane. For the power-law rigidity profiles there is no characteristic spatial scale, so the only parameter of the problem having the dimension of a length is the inverse wave number k^{-1}. Consequently, it is appropriate to normalize the (x, y, z) coordinates of this value: $x' \rightarrow kx, y' \rightarrow ky, z' \rightarrow kz$, and then omit the primes.

The Helmholtz equation for the SH component $u_z = u_z(y)$ of the field (see Fig. 1) can be written independently of the P+SV component in the form:[10,11,18]

$$(c_S^2(y)u_z')' + [(\omega/k)^2 - c_S^2(y)]u_z = 0, \qquad (1)$$

with the normalized on k^{-1} vertical coordinate y, where the primes denote the derivative over y. The equation is subject to the following conditions:

$$c_S^2(y)u_z'(y)|_{y=0} = 0, \quad u_z(y)|_{y\to\infty} = 0, \quad u_z(y)|_{y=0} = 1. \qquad (2)$$

The first boundary condition, Eq. (2), corresponds to the absence of the shear stress σ_{yz} at the surface, while the second one ensures localization of the modes; both are necessary to determine the eigenvalue and the shapes of the eigenfunctions. The third condition $u_z(y=0)=1$ actually represents the choice of normalization for the eigenfunctions.

The Helmholtz equations for the acoustic displacement in the sagittal plane (P+SV waves) are

$$(c_S^2(y)u_x')' + ((\omega/k)^2 - c_L^2(y))u_x$$
$$= i[(c_L^2(y) - 2c_S^2(y))u_y' + (c_S^2(y)u_y)'],$$

$$(c_L^2(y)u_y')' + ((\omega/k)^2 - c_S^2(y))u_y$$
$$= i[c_S^2(y)u_x' + ((c_L^2(y) - 2c_S^2(y))u_x)'], \qquad (3)$$

with the conditions

$$[c_S^2(y)(u_x' - iu_y)]|_{y=0} = 0,$$

$$[c_L^2(y)u_y' - i(c_L^2(y) - 2c_S^2(y))u_x]|_{y=0} = 0,$$

$$u_{x,y}|_{y\to\infty} = 0, \quad u_x|_{y=0} = 1. \qquad (4)$$

Here again the two first boundary conditions correspond to the stress-free boundary with $\sigma_{xy} = \sigma_{yy} = 0$, the third one is the

localization condition, and the last one fixes the normalization of the eigenfunctions. The total number of conditions is five as is necessary for two ordinary differential equations with one unknown eigenvalue.

Following numerous papers[6,8,13,15,16] we assume that the action of gravity on the rigidity of the unconsolidated granular material results in the power-law depth dependence for the longitudinal and the shear sound velocities:

$$c_L(y) = c_{L0}y^\alpha, \quad c_S(y) = c_{S0}y^\alpha, \tag{5}$$

where $c_{L0,S0} = \gamma_{L,S}(\rho g/k)^\alpha$. The granular material is assumed to be arbitrary pliable[27] and the possible anisotropy of the elastic moduli[28] and the material density stratification are neglected for the compactness of the presentation. We also neglect interaction of the elastic waves propagating through the granular skeleton with the compression waves propagating mostly through air-filled pores.

As soon as the forms Eq. (5) for an unconsolidated material are accepted, it becomes obvious how to describe the presence of adhesion between grains[29] including one associated with possible interstitial liquid bridges between the grains.[27,30] Indeed, adhesion can be characterized as tensile stress p_0 that is necessary to apply to a volume of material in order to break the intergrain bounds, in the absence of gravity. This means the actual presence of a hydrostatic pressure p_0. Such pressure can be found in the unconsolidated material at the depth y_0 so that $p_0 = \rho g y_0/k$, for the dimensionless y_0, or, in other words, the state at the surface $y=0$ of the adhesive grains with adhesion pressure p_0 is the same as the state of the unconsolidated grains at the depth y_0. This simple consideration yields the sound velocity profiles for the adhesive material in the form:

$$c_L(y) = c_{L0}(y + y_0)^\alpha, \quad c_S(y) = c_{S0}(y + y_0)^\alpha. \tag{6}$$

Note that for SH waves in the material without adhesion these velocity profiles also correspond to the case where normal pressure p_0 is applied to the surface of a granular material through the loading plate, if the contact between the plate and the granular medium is frictionless. For brevity, we shall call the profiles Eq. (6) "the case with adhesion."

In the absence of adhesion Eqs. (1) and (3) take the forms

$$(y^{2\alpha}u_z')' + (\Omega^2 - y^{2\alpha})u_z = 0 \tag{7}$$

and

$$(y^{2\alpha}u_x')' + (\Omega^2 - \delta y^{2\alpha})u_x = i[(\delta - 2)y^{2\alpha}u_y' + (y^{2\alpha}u_y)'],$$

$$\delta(y^{2\alpha}u_y')' + (\Omega^2 - y^{2\alpha})u_y = i[y^{2\alpha}u_x' + (\delta - 2)(y^{2\alpha}u_x)'], \tag{8}$$

respectively, where the normalized frequency $\Omega = \omega/(c_{S0}k)$ and the parameter $\delta = c_{L0}^2/c_{S0}^2$ are introduced. In these equations there is no spatial parameter, and the length scale of the solution is determined by the penetration depth $\sim k^{-1}$, i.e., depends only on frequency. This means that the solutions of Eqs. (7) and (8) are self-similar. If the grain adhesion is present, the self-similarity is lost. The same happens if the longitudinal and shear velocity profiles do not follow the

power-law dependence with the same α, as for instance for the seismo-acoustic P+SV interface waves in marine sediments[10,11] with the power-law shear velocity profile and almost not stratified longitudinal sound velocity. The violation of the exact proportion $c_L(y)/c_S(y) = \text{const}$ actually means that another spatial scale appears in the system. Such a spatial scale can be introduced also by the influence of the container walls[3,27,31] (due to frictional transfer of the load to the walls, for example), or by crashing the contacts, fragmentation of the beads and consolidation of material at high pressures, or by interaction of the waves propagating through the granular skeleton with compression waves propagating through the air. For grains with adhesion and the corresponding bulk sound velocity profiles, Eq. (6), it is sufficient to substitute y by $y+y_0$ in Eqs. (7) and (8).

It is important to note that the formulated problem for the GSA modes in granular materials without adhesion represents a curious class of problems, in which some boundary conditions are satisfied automatically because of the coefficients in the equations having zero values at the surface. Indeed, the first condition in Eq. (2) and the first and the second ones in Eq. (4) are satisfied since $c_S(0) = c_L(0) = 0$. In this case, to ensure the uniqueness of the solution, as well as from continuity considerations for $y_0 \to 0$, we require for $y_0 = 0$ the fulfillment of the same boundary conditions as for $y_0 \neq 0$, namely

$$u_z'(y)|_{y=0} = 0, \quad u_z(y)|_{y\to\infty} = 0, \quad u_z(y)|_{y=0} = 1 \tag{9}$$

and

$$(u_x' - iu_y)|_{y=0} = 0, \quad [u_y' - i(1 - 2/\delta)u_x]|_{y=0} = 0,$$

$$u_{x,y}|_{y\to\infty} = 0, \quad u_x|_{y=0} = 1 \tag{10}$$

for the SH and P+SV waves, respectively. However, for $y_0 = 0$ the zero surface rigidity makes the solutions insensitive to the exact values of the derivatives u_x', u_y', u_z' at the surface. In fact, it is enough to require that they do not diverge to get solutions very close to those obtained with the exact values of the derivatives according to Eqs. (9) and (10). This feature is confirmed by numerical simulations described in Sec. V and enables us to use a semianalytical approach with Laguerre polynomials (Sec. IV), in which the values of surface derivatives are not controlled. Even high variations of u_x', u_y', u_z' at the surface will affect the profiles $u_{x,y,z}(y)$ only slightly and practically will not change the dispersion relations. Thus, the zero surface rigidity results in a principal difference between the GSA modes in granular materials without adhesion and the Rayleigh surface wave in homogeneous solids.

Later on, depending on the considered case, we shall require the fulfillment of either Eqs. (2) and (4) or Eqs. (9) and (10) having in mind that it is exactly the same for $y_0 \neq 0$ and almost the same for $y_0 = 0$.

III. ANALYTICAL SOLUTIONS FOR THE LOCALIZED SH MODES

When the eigenvalues Ω^2 are found, the form of the dispersion relation can be easily derived by comparing the

Aleshin, Gusev, and Tournat: Acoustic modes in granular media

relationships $c_{S0}=\gamma_S(\rho g/k)^\alpha$ and $\Omega=\omega/(c_{S0}k):\omega=\Omega\gamma_S(\rho g)^\alpha k^{1-\alpha}$. For the velocity $c_{SH}=\omega/k$ of the SH waves that is typically measured in the experiments we immediately get

$$c_{SH}=\Omega\gamma_S(\rho g)^\alpha k^{-\alpha}. \tag{11}$$

In the absence of adhesion, it is obvious that Ω is a k-independent constant, since in Eq. (7) there are no k-dependent quantities. In contrast, if another spatial parameter is introduced, Ω will depend on k through the normalized value of y_0. Anyway, in order to find the proportionality constant Ω in Eq. (11) and acoustic wave shape functions, describing in-depth wave profiles, Eq. (7) should be solved.

A. Geometrical acoustics approximation

A possible approach would be to apply the geometrical acoustics (GA) method (Wentzel, Kramers, Brillouin approximation[10,11,18,19,32]), in which both the amplitude and the phase (eikonal S) of u_z are assumed to be slowly varying at the scale of k_y^{-1}, the latter being itself depth-dependent. For the acoustic rays directed from the surface, the eikonal equation $(S')^2=\Omega^2/y^{2\alpha}-1$ predicts the turning points (simple caustics) at depth $y_c=\Omega^{1/\alpha}$. However, for the power-law profile with $\alpha<1$ the GA validity conditions fail not only as usually in the vicinity of caustic[10,11,18,19] but also in the vicinity of the free surface due to the singularity of the coefficients of the wave equation at $y=0$. So when applying the resonance (quantization) condition of Bohr-Sommerfeld[20,33] to the evaluation of the dispersion relation for the waves guided between the surface and the caustic, it is necessary to take into account not only the phase loss $\Delta\varphi_C$ due to the wave reflection from a simple caustic,[10,11,18] but also the phase loss $\Delta\varphi_0$ corresponding to the reflection from the material rigidity gradient[10,11] $\rho(c_S^2)'$ at $y=0$ (note that this gradient is infinite for $\alpha<1/2$ assumed hereupon):

$$2\int_0^{y_c}dS+\Delta\varphi_c+\Delta\varphi_0=2\pi(n-1), \tag{12}$$

where n is an integer number. The phase loss $\Delta\varphi_C$ is known to be equal[10,11,18] to $-\pi/2$ and $\Delta\varphi_0$ has to be evaluated.

To calculate $\Delta\varphi_0$, we note that in the vicinity of the surface, the incidence of the wave is almost normal, since the depth-dependent vertical component of the wave vector $k_y\to\infty$ due to the vanishing rigidity. This means that the problem can be considered as locally plane, and the phase shift $\Delta\varphi_0$ arising from the surface reflection can be estimated for purely normal incidence. In other words, in the vicinity of the surface $(y\to0)$ in the original Eq. (7) the third term $-y^{2\alpha}u_z$ can be omitted in comparison with Ω^2. The general analytical solution of Eq. (7) without this term is known:[10,11]

$$u_z=\xi^{-\nu}[c_1J_\nu(\xi)+c_2Y_\nu(\xi)]. \tag{13}$$

Here J_ν and Y_ν are the Bessel functions[34] of order $\nu=(\alpha-1/2)/(1-\alpha)$, $\xi\equiv\Omega y^{1-\alpha}/(1-\alpha)$, and $c_{1,2}$ are arbitrary constants. Only the first term in the solution Eq. (13) satisfies stress free boundary condition [Eq. (2)] at the surface

$$u_z=c_1\xi^{-\nu}J_\nu(\xi). \tag{14}$$

The value of $\Delta\varphi_0$ has the sense of the phase shift accumulated after complete reflection of the wave, i.e., it can be found using the asymptotic behavior of the solution Eq. (14) at large ξ:

$$u_z(\xi\gg1)\propto\cos\left(\xi-\nu\frac{\pi}{2}-\frac{\pi}{4}\right)$$
$$\propto\exp\left[i\left(\xi-\nu\frac{\pi}{2}-\frac{\pi}{4}\right)\right]+\exp\left[-i\left(\xi-\nu\frac{\pi}{2}-\frac{\pi}{4}\right)\right].$$

The first exponent here is clearly identified with the wave incident on the surface and the second one with the reflected wave. Correspondingly, the phase shift is equal to $2\xi-\nu\pi-\pi/2$, where the first contribution just accounts for the time of wave propagation to the surface and back, i.e., can be associated to the geometrical phase increment. Consequently, the phase loss $\Delta\varphi_0$ at the discontinuity of material rigidity gradient at the surface $y=0$ is equal to

$$\Delta\varphi_0=-\frac{\pi}{2}(2\nu+1)=-\frac{\pi}{2}\frac{\alpha}{(1-\alpha)}.$$

Note that this theoretical prediction differs from the value of $\Delta\varphi_0$ proposed elsewhere.[10,11]

The first term in Eq. (12), accounting for the geometrical phase increment, is also evaluated analytically. As a result, for the unknown parameter Ω in the dispersion relationship $\omega=\Omega\gamma_S(\rho g)^\alpha k^{1-\alpha}$ for the localized SH waves we have

$$\Omega=\left\{\pi\left[2n-1-\frac{2\alpha-1}{2\alpha-2}\right]\alpha\frac{\Gamma((1/\alpha+2)/2)}{\Gamma((1/\alpha-1)/2)\Gamma(3/2)}\right\}^\alpha$$
$$\equiv(d_n^{GA})^\alpha, \tag{15}$$

where Γ denotes the gamma function.[34] In accordance with Eq. (15) for positive Ω an integer n should satisfy inequality $n>(\alpha-3/4)/(\alpha-1)$, which for $\alpha<1/2$ of our interest here is equivalent to $n>0$. The normal modes in Eq. (15) are numbered in order of increasing phase velocity and, consequently, increasingly deep penetration into the media. Note that in accordance to Eq. (15) $\Omega\propto(n-a)^\alpha$ (with $a\equiv1+[1/(\alpha-1)]/4$) and, since $y_c=\Omega^{1/\alpha}$, for the normalized depth y_c where the caustic is located, we have the proportionality $y_c\propto n-a$, i.e., for the dimensional caustic depth $y_c^*\equiv y_c/k$ we have $y_c^*\propto(n-a)\lambda$. So this caustic depth grows both with mode wavelength λ and the mode order n.

In the important cases of $\alpha=1/4$ and $\alpha=1/6$, Eq. (15) can be rewritten as

$$\Omega=[4/3+4(n-1)]^{1/4},\quad\Omega=[8/5+16(n-1)/3]^{1/6}, \tag{16}$$

respectively. So, the dispersion relations are given by Eq. (11) with the parameter Ω determined by Eq. (15) or Eq. (16).

Although the dispersion relation is derived, it is important to verify that the predicted linear waves do not cause yield (loss by material of elasticity property when the ratio of

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Aleshin, Gusev, and Tournat: Acoustic modes in granular media   2603

shear to normal stress exceeds a critical value of the order of 1). This can be done, for example, by using the solution Eq. (14) near the surface. It shows that the shear acoustic stress σ_{yz} diminishes when $y \to 0$ with the same power law $\sigma_{yz} \propto y$ as the gravity-induced static normal pressure $p \propto y$. So it is always possible to chose the amplitude of the SH wave small enough to be in the regime $|\sigma_{yz}| \ll |\sigma_{yy}| \propto p$ and to avoid yield. This is an important point because, if the yield threshold $|\sigma_{yz}| \propto |\sigma_{yy}|$ is approached, the nonlinear effects such as dilatancy[9,29] might play a crucial role in wave reflection from the surface of granular material and future developments of the theory should take this into account.

It is also worth mentioning that the analytical result in Eq. (14) [and also a more general one in Eq. (19)] demonstrates that for the values $\alpha < 1/2$ of interest here the surface strain is zero $u_z'(y=0)=0$. Consequently, the available analytical results are in favor of replacing in the subsequent numerical analysis the boundary conditions of zero stresses in Eqs. (2) and (4) by the conditions on the surface strains in Eqs. (9) and (10).

The obtained results were related to granular materials without adhesion. It is straightforward to obtain the dispersion relations for the SH modes in the GA approximation in the case of grains with adhesion. For the profiles following Eq. (6) with nonvanishing elastic moduli, it is well known that $\Delta \varphi_0 = 0$. The nonvanishing pressure at the surface introduces in the considered physical system both the spatial scale $y_0^* = p_0/(\rho g)$ in the dimensional variables [or the normalized on k^{-1} value $y_0 = p_0/(\rho g k)$] and the scale for the acoustic velocity $c_0 = c_S(y=0) = c_{S0} y_0^\alpha = \gamma_S p_0^\alpha$. The depth profile of the shear acoustic velocity can be described by $c_S = c_0(1 + y/y_0)^\alpha$ [Eq. (6)]. This has been taken into account when evaluating the geometrical phase increment and finally the pressure-dependent dispersion relations for the SH waves velocity c_{SH}:

$$\frac{1}{\alpha}\left(\frac{c_{SH}}{c_0}\right)\left\{\frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{\alpha^{-1}-1}{2}\right)}{\Gamma\left(\frac{\alpha^{-1}+2}{2}\right)}\left(\frac{c_{SH}}{c_0}\right)^{1/\alpha-1}\right.$$

$$\left.-\frac{2\alpha}{1-\alpha}F\left[-\frac{1}{2},\frac{1}{2}(\alpha^{-1}-1);\frac{1}{2}(\alpha^{-1}-1);\left(\frac{c_{SH}}{c_0}\right)^{-2}\right]\right\}$$

$$= 2\pi(n-3/4)/(ky_0^*) \equiv 2\pi(n-3/4)/q. \qquad (17)$$

Here F denotes Gauss hypergeometric function.[34] The dependence of normalized phase velocity c_{SH}/c_0 on the normalized wave number $q \equiv ky_0^*$ of the SH modes given implicitly by this solution is presented in Fig. 2(a) for $\alpha = 1/4$ and $\alpha = 1/6$. For long acoustic wave with $q \ll 1$ the phase velocity increases infinitely in the absence of saturation of material moduli growth with depth, and the self-similar asymptotic behavior $c_{SH}/c_0 \propto q^{-\alpha}$ [see Eq. (11)] shows up [Fig. 2(b)], but the short acoustic waves with
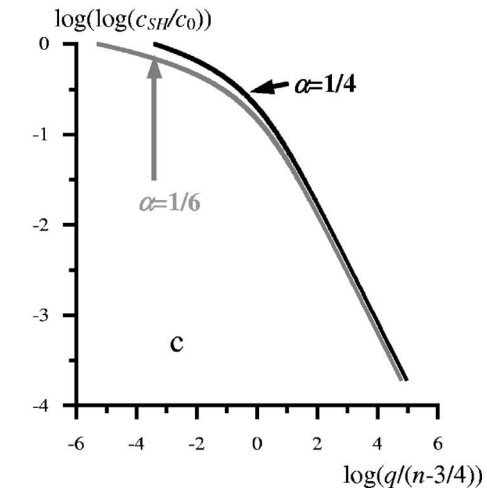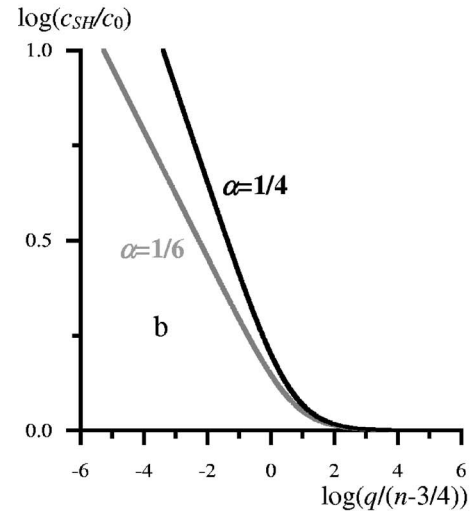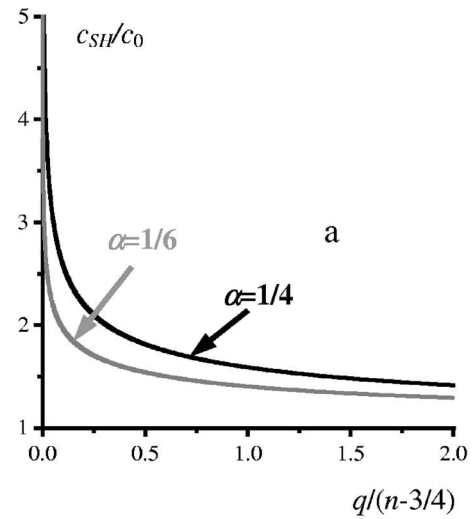
FIG. 2. Dependence of the normalized phase velocity c_{SH}/c_0 on the normalized wave number q for the SH modes, for $\alpha = 1/4$ and $\alpha = 1/6$ (a). The long wave limit of this dependence $q \ll 1$ (b) and the short wave $q \gg 1$ limit (c).

$q \gg 1$ feel the perturbation (truncation) of the power-law shear velocity profile by pressure p_0. In log-log plot in Fig. 2(b) this manifests as a deviation of the curves from the straight lines presenting the self-similar limit. The velocity of the GSAM approaches the characteristic velocity c_0 also following the power law $c_{SH}/c_0 - 1$

$\approx \frac{1}{2}(3\pi\alpha)^{2/3}[q/(n-3/4)]^{-2/3}$. This is illustrated in Fig. 2(c). The dependence of the SH localized modes on pressure arrives in the obtained solution from $c_0 = \gamma_S p_0^\alpha$ and $y_0^* = p_0/(\rho g)$.

The accuracy of the GA approximation is known to increase for higher order modes with increasing n and to be the worse for the lowest mode[10,11,18] $n=1$. In the considered physical system the deficiency of the GA approach additionally manifests itself in the abrupt variation of the phase loss $\Delta\varphi_0$ in the transition between the regimes $p(y=0)=0$ and $p(y=0)\neq 0$. In the GA (short wavelength approximation) the phase loss is an exclusive feature of the surface delta-localized singularity of the material elastic properties, and it abruptly changes when the power-law type singularity is replaced by step-function type singularity. Thus for the quantitative analysis of the lowest modes the analytical solutions of Eq. (7) are highly desirable.

B. Exact analytical solution of the Helmholtz equation in a particular case

Until now the analytical solutions of Eq. (7) have been known[10,11,18] only for the powers -1, 0, $\frac{1}{2}$, and 1. In particular, for $\alpha=1/2$ and $\alpha=1$ the equation for the SH waves can be transformed in the form of the stationary Schrödinger equation for the wave function in a potential, which includes the terms proportional to $1/y^2$ and $1/y$ or $1/y^2$ only. In quantum mechanics these equations are suitable for the description of charge motion in the Coulomb potential,[33] but they are not relevant to unconsolidated granular materials.

We have managed for the first time to solve Eq. (7) in the important for the granular materials case $\alpha=1/4$ by transforming it into the Kummer's equation for the confluent hypergeometric functions. The previously unknown general analytical solution is

$$u_z(y) = e^{-\eta}\left[c_1 U\left(a,\tfrac{1}{2},2\eta\right) + c_2 M\left(a,\tfrac{1}{2},2\eta\right)\right], \quad (18)$$

where U and M are the Kummer functions, $\eta = \eta(y) = [y^{1/2} - d^{1/2}/2]^2$, $a = [1-d/2]/4$, and $d=\Omega^4$. The case of granular material with adhesion, characterized by the parameter y_0 can be easily taken into account by taking $\eta = \eta(y+y_0) = [(y +y_0)^{1/2} - d^{1/2}/2]^2$. Note that the previously known analytical solutions for the powers -1, 0, $\frac{1}{2}$, and 1 in Eq. (7) had been also obtained[8,33] by reducing it to the confluent hypergeometric equation, but for the Whittaker function instead of the Kummer functions.

The items in the solution (18) have the following essential properties: Both have an infinite derivative at $y=0$, the U component decays at $y\to\infty$ while the modulus of M component infinitely grows which is plotted in Fig. 3(a) illustrating a particular example with $d=4$. In addition, the U component
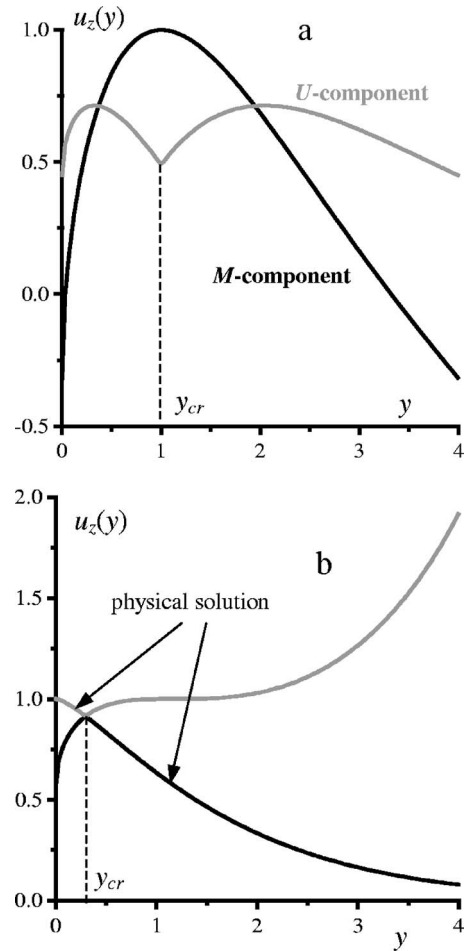
FIG. 3. (a) The components $e^{-\eta}M(a,\tfrac{1}{2},2\eta)$ (black) and $e^{-\eta}U(a,\tfrac{1}{2},2\eta)$ (gray) of the general solution $u_z(y)$ [Eq. (18)] plotted for $d=4$. Both components have an infinite derivative at $y=0$. The U component decays at $y \to\infty$ and has a discontinuous derivative at $y_{cr}=d/4$. The modulus of M component infinitely grows at $y\to\infty$. (b) The shape of the exact solution Eqs. (19) and (20) for the first mode ($d=1.1802$) that follows the gray curve [the first line in Eq. (19)] $y<y_{cr}$ and the black one for $y>y_{cr}$. This solution has the continuous derivative at $y=y_{cr}$.

has a discontinuous derivative at $y_{cr}=d/4$ [or $y_{cr}=d/4-y_0$ for nonzero y_0]. This means that the range $0<y<\infty$ has to be divided into two regions by the internal point y_{cr}, so that three following requirements are met:

(1) For $y<y_{cr}$ both components of the solution are present with some coefficients c_1 and c_2, in order to have a possibility of obtaining zero derivative $u_z'(0)=0$, which is required by the first boundary condition, Eq. (9).
(2) For $y>y_{cr}$, only the decaying U component is present.
(3) At the critical point $y=y_{cr}$ the solution is smooth, i.e., it is continuous itself and has a continuous derivative.

These conditions are possible to satisfy by taking

$$u_z(y) = ce^{-\eta}\begin{cases} 2\pi^{1/2}M\left(a,\tfrac{1}{2},2\eta\right)/\Gamma\left(a+\tfrac{1}{2}\right) - U\left(a,\tfrac{1}{2},2\eta\right) & \text{for } 0 \leq y \leq y_{cr} \\ U\left(a,\tfrac{1}{2},2\eta\right) & \text{for } y \geq y_{cr}, \end{cases} \quad (19)$$
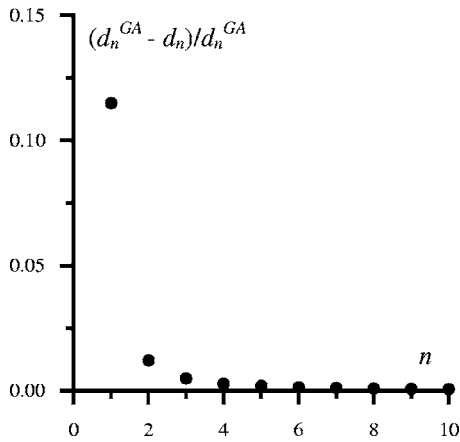
FIG. 4. The accuracy of the GA approximation for the SH modes. The relative deviation of the dispersion parameter d_n^{GA} from the exact value d_n of this parameter is plotted as a function of the mode number.

FIG. 5. The depth profiles of the displacement for the five lowest SH modes in the absence of adhesion ($y_0=0$). The number of zeros for each mode corresponds to the order of the mode.

where $y_{cr}=d/4$. It is straightforward to verify analytically that this form of the solution ensures its smoothness and fulfillment of the localization condition [see Eq. (9)]. The constant c is arbitrary and can be fixed so that $u_z(0)=1$ according to Eq. (9). The boundary condition $u_z'(0)=0$ corresponding to the absence of the shear stress $\sigma_{yz}=0$ can be satisfied by adjusting the free parameter d which leads to the algebraic equation

$$\pi^{1/2}\left[4a_n M\left(a_n+1,\tfrac{3}{2},2\eta_n\right)-M\left(a_n,\tfrac{3}{2},2\eta_n\right)\right]/\Gamma\left(a_n+\tfrac{1}{2}\right)$$
$$+a_n U\left(a_n+1,\tfrac{3}{2},2\eta_n\right)+\tfrac{1}{2}U(a_n,b,2\eta_n)=0, \qquad (20)$$

where $\eta_n=(0-d_n^{1/2}/2)^2=d_n/4$, $a_n=[1-d_n/2]/4$, and n numbers the roots d_n of this equation. The presence of adhesion with the parameter y_0 can be easily taken into account by substituting $\eta_n=(y_0^{1/2}-d_n^{1/2}/2)^2$ and $y_{cr}=d/4-y_0$ in Eqs. (19) and (20). The roots d_n contribute directly to the dispersion relation Eq. (11) through $d=\Omega^4$.

The two branches of the solution Eqs. (19) and (20) are plotted in Fig. 3(b), for $d=d_1\approx1.180$ corresponding to the first mode, in the case without adhesion. The gray curve corresponds to the first line in Eq. (19), the black curve represents the second line with the U function only. As it was mentioned, at $y=y_{cr}=d_1/4$ the solution has a continuous derivative.

As it has been expected, the dispersion parameters d_n obtained through the solution of Eq. (20) in the case $y_0=0$ importantly differ from the prediction $d_n^{GA}=4/3-4(n-1)$ of the GA approximation [Eq. (16)] only for the lowest modes (see Fig. 4). The sequence $\{d_n\}$ is not equally spaced: $d_{n+1}-d_n\neq d_n-d_{n-1}$, in contrast to the equidistant $\{d_n^{GA}\}$, but rapidly becomes so when n increases. The displacements fields for the five lowest eigenmodes in the limit of negligible adhesion are presented in Fig. 5. We can also see that the order n of the mode can be identified with the number of its profile zeros including the one at $y\rightarrow\infty$, or with the number of the "phases" in the depth profile. Thus, for the first mode, the displacements of the particles at all depths are in phase (single-polar or single-phase depth profile). The profile of the
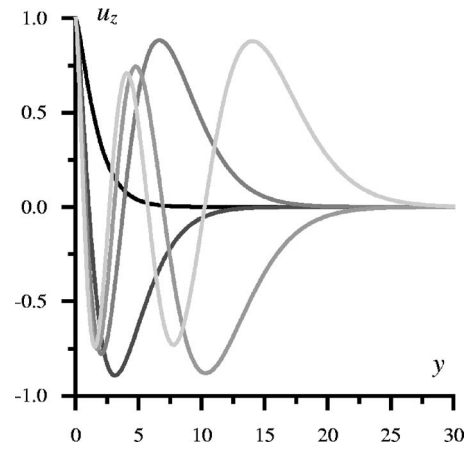
second mode is bipolar, i.e., it contains two consecutive phases. The third mode profile contains three alternating phases and so on.

With the increasing pressure $p_0\neq0$, the matching surface $y=y_{cr}=d_n/4-y_0$ approaches the materials surface, since $y_0\propto p_0$. It can be shown that the description of the displacement field u_z [Eq. (19)] includes both Kummer functions in Eq. (7) only when the phase velocity of SH acoustic modes satisfies the inequality $c_{SH}>2^{1/2}c_0$. The displacement field for the slower modes is described in the entire half space $0\leq y\leq\infty$ just by the second line of Eq. (19). Correspondingly, in Eq. (20) for the dispersion parameter d_n, the Kummer function M should be omitted. In Fig. 6 the dependence of d_n on $y_0\propto p_0$ is illustrated: The increase in pressure contributes both to the increase in the wave velocity, and to its additional relatively to $c_{SH}\propto k^{-\alpha}$ dispersion. The wave depth profiles (Fig. 7) also start to be pressure dependent: With increasing pressure, the modes penetrate deeper into the medium, and the self-similarity of the in-depth profiles is lost. Finally, for a very large y_0, when the medium rigidity is almost homogeneous, the modes delocalize. This is in accordance with
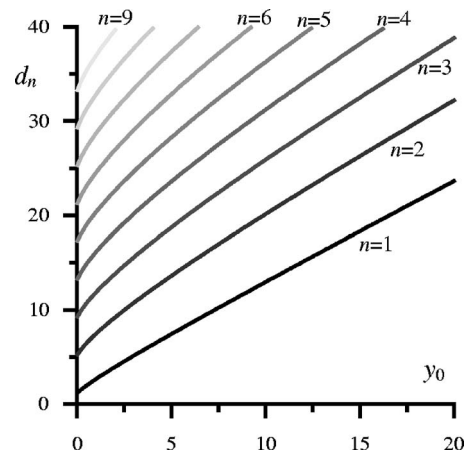
FIG. 6. The dependence of the dispersion parameters d_n on the adhesion parameter y_0 proportional to p_0. The values of d_n contribute directly to the dispersion relation Eq. (11) via $d_n=\Omega^4$.

Aleshin, Gusev, and Tournat: Acoustic modes in granular media
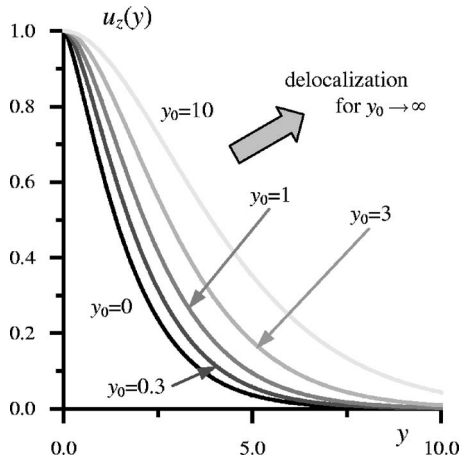
FIG. 7. The depth profiles of the first SH mode for different adhesion parameters y_0 proportional to the pressure $p_0 \to \infty$.

the fact that the only localized mode in a homogeneous medium is the Rayleigh wave with sagittal polarization, and no modes with horizontal polarization exist.

The obtained exact, explicit analytical solution for $\alpha = 1/4$ is quite important. It might be the basis for the perturbation analysis of the distributions with different from $\frac{1}{4}$ but close values of the parameter α, encountered in unconsolidated granular materials. An example of such analysis based on the exact solution for $\alpha = 1/2$ is known.[10,11] This solution can also be matched to the GA solution at larger depths, where the power law with a single power is not valid any more, and the elasticity growth is saturated at some critical depth, e.g., the so-called truncated power-law profile.[10,11] In this case, the applicability of the GA increases strongly due to slower variations in material properties.

IV. SEMIANALYTICAL SOLUTIONS FOR THE SH AND P+SV MODES IN THE CASE OF NEGLIGIBLE ADHESION

The propagation of longitudinal (P) and shear vertical (SV) waves is decoupled only in the high-frequency limit (GA approximation) and only in the case $p_0 = 0$ of zero rigidity at the surface. In the GA approximation, the waves are assumed to be insensitive to the small gradients of elasticity in the bulk material and, consequently, there is no mode conversion under the surface in this approximation. At the same time, the incidence of the rays on the surface $y = 0$ with zero rigidity is nearly normal and, consequently, there will be no mode conversion in reflection exactly at the surface by symmetry principles. Because of this, the solutions for the independently guided P and SV GSA modes can be obtained in GA approximation similarly to what have been proposed earlier in this paper for the SH modes. This analysis is not presented here. Apparently, for the analysis of this limiting case the independent equations for the scalar and vector potential[2] are more suitable than Eq. (3).

In the following we present a semianalytical method valid for arbitrary frequencies. In the high-frequency limit this method provides the opportunity in the case $p_0 = 0$ to precisely describe the lower-order P+SV modes, what is not possible in GA approximation.

For an arbitrary power-law vertical distribution of the elastic moduli, the dispersion relations and the wave field distributions both for SH and P+SV modes can be obtained by expanding the displacement fields in series of the Laguerre polynomials[34] L_n, which are providing a complete system of orthogonal functions for a half-space. The presentation $u_z = a_m \exp(-\varsigma/2)L_m(\varsigma)$, where $0 \leqslant m \leqslant \infty$ is the summation index, has been originally proposed[35] for the SH modes in media with rigidity profiles approximated by exponential functions. Such series suits well the considered problem since the zero surface rigidity makes the solution insensitive to the displacement derivatives at the surface, which are not known *a priori*, and the corresponding stress-free boundary conditions [Eqs. (2) and (4)] are satisfied automatically without any restrictions on the coefficients of the series. The only requirement is that the coefficients found from the solution of the equations of motion, Eqs. (1) and (3), should be finite in magnitude.

The dispersion relations for the GSA modes are obtained through the compatibility condition for the system of equations derived for the coefficients when the series are substituted in Eq. (1) or Eq. (3), multiplied by $y \exp(-y/2)L_n(y)$ and then integrated over $0 \leqslant y \leqslant \infty$. We have verified that the application of this method to Eq. (1) for the SH waves reproduces all the above presented results in the case $p_0 = 0$. The key point of the method is that the overlap integrals of the type $\int_0^\infty \exp(-y)y^\beta L_m(y)L_n(y)dy$ determining the coefficients of the algebraic system of equations for a_m can be for $\beta > -1$ ($\alpha > 0$) evaluated analytically in terms of the generalized hypergeometric function[34] $_3F_2$,

$$I_{m,n}(\alpha) \equiv \int_0^\infty e^{-y}y^{\alpha-1}L_m(y)L_n(y)dy$$

$$= \frac{\Gamma(m+1-\alpha)\Gamma(\alpha)}{\Gamma(1-\alpha)n!} {}_3F_2(-m,\alpha,\alpha;1,\alpha-n;1).$$

Equation (1) for the SH waves is then transformed into the algebraic system $A_{nm}a_m = 0$, $n = 0,1,2,\ldots$ with the coefficients of the matrix A_{nm} given by

$$A_{nm}(\alpha,d) = m\lfloor(m+2\alpha-1)I_{m,n}(2\alpha) - 2(m+\alpha-1)$$
$$\times I_{m-1,n}(2\alpha) + (m-1)I_{m-2,n}(2\alpha)\rfloor - (m+\alpha)$$
$$\times I_{m,n}(2\alpha+1) + mI_{m-1,n}(2\alpha+1) + d^{2\alpha}I_{m,n}(2)$$
$$- (3/4)I_{m,n}(2\alpha+2).$$

The dispersion parameters d connected to the normalized frequency Ω via $\Omega = d^\alpha$ are found as the roots of the equation $\text{Det}[A_{nm}(\alpha,d)] = 0$. The rank of the determinant, i.e., number of Laguerre polynomials in the series, was chosen in order to ensure the desired precision, which is easy to provide as we monitor the saturation of the roots as a function of the cut-off number.

For the waveguide P+SV modes, we search the solution in the form $u_x = b_m \exp(-y/2)L_m(y)$, $u_y = c_m \exp(-y/2)L_m(y)$ and get the matrix equations

$$A_{nm}^x b_m = B_{nm}^y c_m, \quad A_{nm}^y c_m = B_{nm}^x b_m, \quad n = 0,1,2,\ldots,$$

where
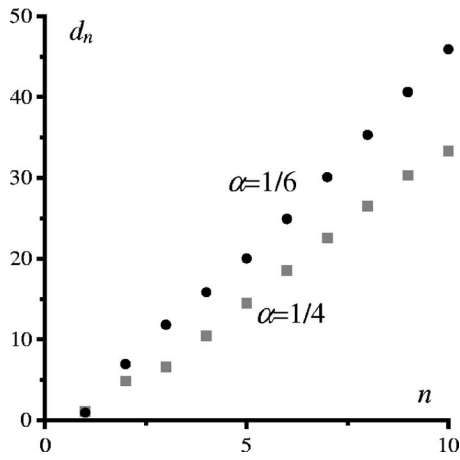
FIG. 8. The dispersion parameters d_n for the P+SV modes. The first ten modes are shown for $\alpha=1/6$ and $\alpha=1/4$. Here $\delta=3$.

$$A_{nm}^x(\alpha,d,\delta) = m[(m+2\alpha-1)I_{m,n}(2\alpha) - 2(m+\alpha-1)$$
$$\times I_{m-1,n}(2\alpha) + (m-1)I_{m-2,n}(2\alpha)]$$
$$-(m+\alpha)I_{m,n}(2\alpha+1) + mI_{m-1,n}(2\alpha+1)$$
$$+d^{2\alpha}I_{m,n}(2) - (\delta-1/4)I_{m,n}(2\alpha+2)$$
$$= A_{nm}(\alpha,d) - (\delta-1)I_{m,n}(2\alpha+2),$$

$$B_{nm}^y(\alpha,d,\delta) = i\{[2\alpha + m(\delta-1)]I_{m,n}(2\alpha+1)$$
$$- m(\delta-1)I_{m-1,n}(2\alpha+1) - (1/2)(\delta-1)$$
$$\times I_{m,n}(2\alpha+2)\},$$

$$A_{nm}^y(\alpha,d,\delta) = \delta\{m[(m+2\alpha-1)I_{m,n}(2\alpha) - 2(m+\alpha-1)$$
$$\times I_{m-1,n}(2\alpha) + (m-1)I_{m-2,n}(2\alpha)]$$
$$-(m+\alpha)I_{m,n}(2\alpha+1) + mI_{m-1,n}(2\alpha+1)\}$$
$$+d^{2\alpha}I_{m,n}(2) - (1-\delta/4)I_{m,n}(2\alpha+2)$$
$$= \delta A_{nm}(\alpha,d) - (\delta-1)[d^{2\alpha}I_{m,n}(2)$$
$$- I_{m,n}(2\alpha+2)],$$

and

$$B_{nm}^x(\alpha,d,\delta) = i\{[2\alpha(\delta-2) + m(\delta-1)]I_{m,n}(2\alpha+1)$$
$$- m(\delta-1)I_{m-1,n}(2\alpha+1)$$
$$- (1/2)(\delta-1)I_{m,n}(2\alpha+2)\}.$$

Here the presentation of $A_{nm}^{x,y}$ matrices is given in the compact form using the previously introduced A_{nm} matrix for the SH modes. It is easy to get rid of the imaginary unit i by substituting $iu_y \to u_y$. The dispersion parameters d are found as the roots of the equation $\mathrm{Det}|A_{lm}^x - B_{ln}^y(A_{kn}^y)^{-1}B_{km}^x|=0$. The ten first roots are plotted in Fig. 8 for a particular value of the ratio $\delta=c_{L0}^2/c_{S0}^2$, while the dependence of the first four roots on δ is shown in Fig. 9.

To compare the theoretical predictions with the available measurement of the velocity of the P+SV mode[24] we present the derived solution in the dimensional form $c_{P+SV}^{theor} = \gamma_S(\rho g d_n/k)^\alpha$ [see Eq. (11)] and compare it with the experimental data on the bulk shear acoustic velocity dependence
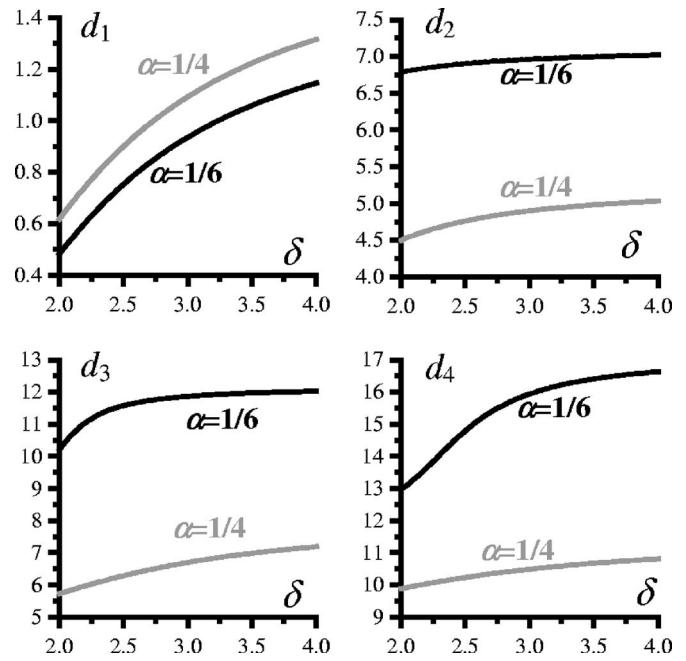
FIG. 9. The dispersion parameters d_n as functions of $\delta=c_{L0}^2/c_{S0}^2$, for $\alpha=1/4$ and $\alpha=1/6$.

on pressure $c_S^{exper} = \gamma_S p_{exper}^\alpha$. We derive $c_{P+SV}^{theor} = c_S^{exper}[\rho g d_n/(p_{exper}k)]^\alpha$. With $\rho \approx 1560$ kg/m^3, $c_S^{exper} = 800$ m/s, $c_L^{exper} = 1400$ m/s, $\delta \approx 3$, $\alpha \cong 1/4$ at $p_{exper} \approx 20$ MPa from the published data[6,15] for $k \approx 20\pi$ m^{-1} ($\lambda \approx 0.1$ m) from the experimental range[24] it has been found $c_{P+SV}^{theor} \approx 45 d_n^{1/4}$ m/s. Thus the theoretical prediction for the lowest mode without free parameters differs only about 10% from the experimental value $c_{P+SV}^{exper}(\lambda \approx 0.1$ m$) \approx 40$ m/s. The persistent discrepancy between the theoretical prediction and the experimental data could be attributed to many factors, among them the difference of sand in Sahara[24] from the glass beads in experiment[6] as well as from the Ottawa sand in experiment[15] (in particular, the difference in humidity of air saturating the sand), and to an uncertainty of the experimental data.

V. NUMERICAL SOLUTION OF THE HELMHOLTZ EQUATIONS FOR THE SH AND P+SV MODES

In this theoretical study, we preferred to build up analytical methods and to restrict ourselves from numerical analysis of the problem. However, such a numerical solution is easy to obtain, especially for the SH modes. In this case, we numerically solve Eq. (1) as a boundary problem with the second and the third boundary conditions in Eq. (9), for a set of arbitrary $\Omega=d^\alpha$, and then adjust d in order to satisfy the first boundary condition $u_z'(0)=0$. The adjustment can be realized by standard eigenproblems methods, e.g., taking an interval between two neighboring points of the trial array of d so that the derivative $u_z'(0)$ changes its sign in this interval, and then localize the root by dichotomy (bisectional method) with any desired precision.

For the set Eq. (3) and boundary conditions, Eq. (10), the situation is slightly more complicated. First of all, as

Aleshin, Gusev, and Tournat: Acoustic modes in granular media

previously, we substitute iu_y by u_y to get rid of the imaginary unit i. Then the eigenproblem is equivalent to the boundary problem with boundary conditions

$$u_x(0) = 1, \quad u_x(\infty) = 0, \quad u_y(0) = \gamma, \quad u_y(\infty) = 0, \qquad (21)$$

where γ is an unknown parameter that should be adjusted together with d in order to fulfill the conditions:

$$u'_x = \gamma, \quad u'_y = (2 - \delta)/\delta. \qquad (22)$$

A double parameter matching is numerically expensive; this difficulty can be avoided by consideration of the two following auxiliary problems. The first one is for functions $v_x(y)$ and $v_y(y)$ satisfying Eq. (3) and the conditions

$$v_x(0) = 1, \quad v_x(\infty) = 0, \quad v_y(0) = 0, \quad v_y(\infty) = 0,$$

while the second one for the functions $w_x(y)$ and $w_y(y)$ has the conditions

$$w_x(0) = 0, \quad w_x(\infty) = 0, \quad w_y(0) = 1, \quad w_y(\infty) = 0.$$

For a given d, these auxiliary problems can be solved independently of the unknown γ and yield the derivatives $v'_x(0), v'_y(0), w'_x(0), w'_y(0)$. Since $u_{x,y} = v_{x,y} + \gamma w_{x,y}$, the conditions [Eq. (22)] are transformed into

$$v'_x(0) + \gamma w'_x(0) = 1, \quad v'_y(0) + \gamma w'_y(0) = (2 - \delta)/\delta,$$

or, excluding γ,

$$w'_x(0)[v'_y(0) - (2 - \delta)/\delta] - w'_y(0)[v'_y(0) + 1] = 0. \qquad (23)$$

This procedure reduces the two-parameter adjustment to a one-parameter fit: Varying the only value of d, we find roots of Eq. (23) by the dichotomy method, then calculate $\gamma = (1 - v'_x(0))/w'_x(0)$ and finally build up $u_{x,y} = v_{x,y} + \gamma w_{x,y}$.

The dispersion relations for the SH and P+SV modes, as well as the displacement profiles for the SH modes, have been already given in Secs. III and IV, and they totally agree with the numerical computations. Here we add only displacement profiles for the P+SV waves obtained numerically [Figs. 10(a) and 10(b)]. Using the numerical method it is easy to check sensitivity of the solutions to the values of the derivatives u'_x, u'_y, u'_z at the surface of the material without adhesion. Doing so, we found that, for instance, changing for the SH problem $u'_z(0) = 0$ by $u'_z(0) = 100$ introduces an error of 0.5% only, while the dispersion parameters d_n are practically not affected.

The numerical method works well in the absence and presence of adhesion. In the presence of strong adhesion the medium becomes almost homogeneous near the surface resulting in the corresponding behavior of the modes: The first mode tends to the Rayleigh mode,[1,18] which is the only surface wave that exists in homogeneous solids, the other modes delocalize and lose more and more of their surface character as the adhesion parameter increases. This is illustrated in Figs. 11(a)–11(d) where the first and second modes are plotted against depth y for different values of y_0. For $y_0 = 100$ the difference between the first GSAM and the Rayleigh wave is indiscernible in the plot, while the second mode has a drastically increased penetration depth.

So, we have demonstrated that in the two different limits of absence of adhesion and strong adhesion, the mechanisms
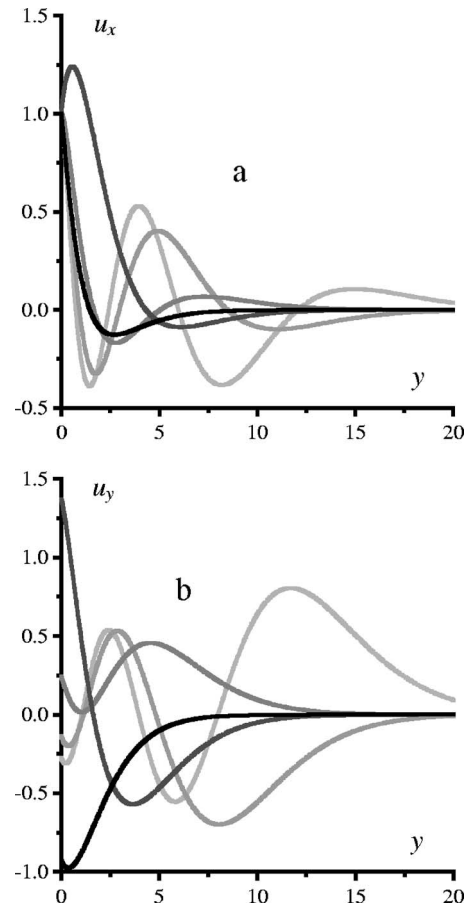
FIG. 10. The depth profiles of the displacement for the five lowest P+SV modes in the absence of adhesion ($y_0 = 0$): u_x component (a) and u_y component (b). $\alpha = 1/4$, $\delta = 3$.

of localization are physically different. For strongly adhesive materials, the localized modes can exist only due to the coupling of P and SV waves on the surface, similar to Rayleigh waves in a homogeneous half-space. The exact fulfillment of the Rayleigh-type boundary conditions on the surface is required. In the absence of adhesion, in contrast, the waves localize not due to the surface coupling of modes, but because of rigidity stratification. Accordingly, there is an infinite number of surface modes for both P+SV and SH polarizations, while in the homogeneous solids only one surface P+SV mode exists. Actually because of this we prefer to use the term GSAM for these modes and to emphasize their waveguide character, rather then to call them Rayleigh waves in granular materials.[24] The GSAMs differ importantly from the SH and Lamb waveguide modes in plates as well, as the dispersion of the latter is introduced entirely by the plate thickness. The dispersion of Love waves, which are SH localized modes propagating along the plane interface between an homogeneous semi-infinite substrate and an homogeneous layer of lower shear velocity than that of the substrate, is also introduced entirely by the layer thickness.

VI. CONCLUSIONS

The channel for the acoustic waves propagation in granular materials analyzed here is one more example of the naturally existing waveguides on the Earth, in addition to

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Aleshin, Gusev, and Tournat: Acoustic modes in granular media   2609
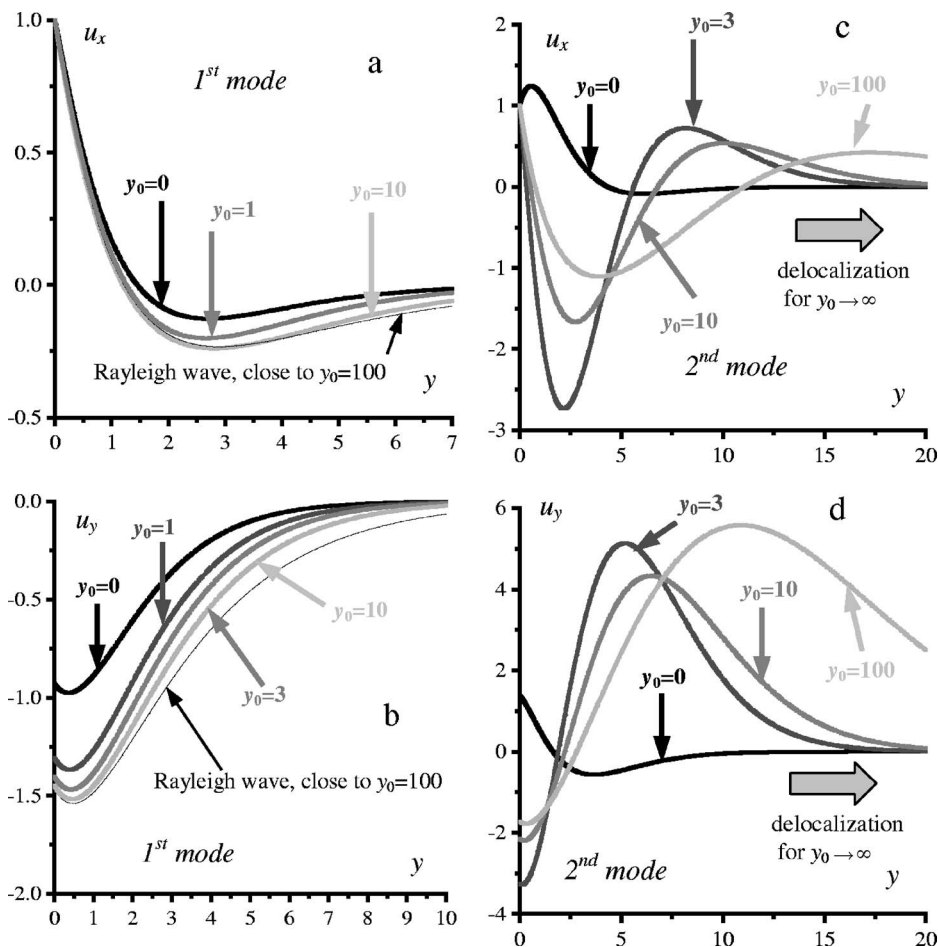
FIG. 11. The depth y profiles of the displacements u_x and u_y for increasing adhesion parameter y_0. The first mode [(a), (b)] tends to the Rayleigh wave: The Rayleigh wave and the profiles for $y_0=100$ are indiscernible in the plot. The second [(c), (d)] and higher modes (not shown) delocalize. The modes are plotted for $\delta=3$ and $\alpha=1/4$.

deep sound channel in the ocean (SOFAR, i.e., sound fixing and ranging channel[36]), channel for the radiowaves between the Earth surface and the ionosphere, channels for seismic waves propagation through the interior of the Earth,[37] etc.

The obtained results might be useful for the inversion of the experimental data, i.e., in order to retrieve information on the vertical stratification from measurements of the dispersion relations of the localized modes.

In practice, the power scaling predicted for the power-law velocity profiles Eq. (5) in the absence of adhesion ($y_0=0$) can be valid for a finite interval of the acoustic wave numbers (wavelengths or frequencies) only. At low frequencies the scaling law breaks since the long-wavelength modes penetrate to the depths where elastic moduli deviate from the power-law profile. At high frequencies the very concept of an effective solid underlying the developed theory does not apply, when the acoustic wavelength becomes comparable to the material grain size. The adhesion of grains, as we have seen earlier, also causes truncation of the power-law profiles.

From both theoretical and experimental points of view, it is interesting to answer the question if there exist nonpropagating (evanescent) waveguide modes[38] localized near the surface of unconsolidated granular material similar to those existing both for SH and Lamb modes in plates. The knowledge of the modes with the nonpurely real values of the propagation constant (evanescent modes) could be important for the description of the acoustic field in a vicinity of a localized transient source of acoustic waves. Another intrigu-

ing direction of future studies can be the influence of the material nonlinearity on the acoustic modes and on the coupling between the SH and P+SV modes. Finally, the physics of the GSA modes can be substantially different for very high frequencies, when the acoustic wavelength becomes comparable to the material grain size or to a characteristic scale introduced in the granular medium by the presence of force chains.[6–8,39,40]

ACKNOWLEDGMENTS

[1]V. V. Krylov, "Effect of surface phenomena in solids on surface acoustic waves," Prog. Surf. Sci. **32**, 39–110 (1989).
[2]S. V. Biryukov, Yu. V. Gulyaev, V. V. Krylov, and V. P. Plessky, *Surface Acoustic Waves in Inhomogeneous Media* (Springer, Berlin, 1995).
[3]J. Duran, *Sables, Poudres et Grains* (Eyrolles, Paris, 1997).
[4]H. Elbelrhiti, P. Claudin, and B. Andreotti, "Field evidence for surface-wave-induced instability of sand dunes," Nature (London) **437**, 720–723 (2005).
[5]R. A. Bagnold, *The Physics of Blown Sand and Desert Dunes* (Methuen, London, 1941).
[6]H. A. Makse, N. Gland, D. L. Johnson, and L. Schwartz, "Granular packings: Nonlinear elasticity, sound propagation, and collective relaxation dynamics," Phys. Rev. E **70**, 061302–061318 (2004).
[7]L. E. Silbert, A. J. Liu, and S. R. Nagel, "Vibrations and diverging length

scales near the unjamming transition," Phys. Rev. Lett. **95**, 098301 (2005).

[8]E. Somfai, J.-N. Roux, J. H. Snoeijer, M. van Hecke, and W. van Saarloos, "Elastic wave propagation in confined granular systems," Phys. Rev. E **72**, 021301 (2005).

[9]V. Tournat, V. Zaitsev, V. Gusev, V. Nazarov, P. Béquin, and B. Castagnède, "Probing weak forces in granular media through nonlinear dynamic dilatancy: Clapping contacts and polarization anisotropy," Phys. Rev. Lett. **92**, 085502 (2004).

[10]O. A. Godin and D. M. F. Chapman, "Shear-speed gradients and ocean seismo-acoustic noise resonances," J. Acoust. Soc. Am. **106**, 2367–2382 (1999).

[11]O. A. Godin and D. M. F. Chapman, "Dispersion of interface waves in sediments with power-law shear speed profiles. I. Exact and approximate analytical results," J. Acoust. Soc. Am. **110**, 1890–1907 (2001).

[12]D. M. F. Chapman and O. A. Godin, "Dispersion of interface waves in sediments with power-law shear speed profiles. II. Experimental observations and seismo-acoustic inversions," J. Acoust. Soc. Am. **110**, 1908–1916 (2001).

[13]J. D. Goddard, "Nonlinear elasticity and pressure dependent wave speed in granular media," Proc. R. Soc. London, Ser. A **430**, 105–131 (1990).

[14]F. Gassmann, "Elastic waves through a packing of spheres," Geophysics **16**, 673–685 (1951).

[15]S. N. Domenico, "Elastic properties of unconsolidated porous sand reservoirs," Geophysics **42**, 1339–1366 (1977).

[16]B. Velicky and C. Caroli, "Pressure dependence of the sound velocity in a two-dimensional lattice of Hertz-Mindlin balls: Mean-field description," Phys. Rev. E **65**, 021307 (2002).

[17]C-H. Liu and S. R. Nagel, "Sound in sand," Phys. Rev. Lett. **68**, 2301–2304 (1992).

[18]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media I: Plane and Quasi-Plane Waves* (Springer, Berlin, 1990).

[19]Yu. A. Kravtsov and Yu. I. Orlov, *Geometrical Optics of Inhomogeneous Media* (Springer, Berlin, 1990).

[20]V. V. Krylov, "New type of vibration dampers utilising the effect of acoustic 'black holes," Acust. Acta Acust. 90, 830–837 (2004).

[21]M. A. Mironov, "Propagation of a flexural wave in a plate whose thickness decreases smoothly to zero in a finite interval," Sov. Phys. Acoust. **34**, 318–319 (1988); Akust. Zh. **34**, 546–547 (1988).

[22]P. H. Brownell, "Compressional and surface waves in sand: Used by desert scorpions to locate prey," Science **197**, 479–482 (1977).

[23]P. H. Brownell, "Prey detection by the sand scorpion," Sci. Am. **251**, 94–105 (1984).

[24]B. Andreotti, "The song of dunes as a wave-particle mode locking," Phys. Rev. Lett. **93**, 238001 (2004).

[25]V. E. Gusev, V. Aleshin, and V. Tournat, "Acoustic waves in an elastic channel near the free surface of granular media," Phys. Rev. Lett. **96**, 214301 (2006).

[26]F. Melo, P. Umbanhowar, and H. L. Swinney, "Transition to parametric wave patterns in a vertically oscillated granular layer," Phys. Rev. Lett. **72**, 172–175 (1994).

[27]Y. Jiang and H. Liu, "Energetic instability unjams sand and suspension," Phys. Rev. Lett. **93**, 148001 (2004).

[28]X. Zheng and B. Ni, "Anisotropic elastic shear stiffness of sands and its measurement," J. Geotech. Geoenviron. Eng. **125**, 741–749 (1999).

[29]P. G. de Gennes, "Granular matter: A tentative view," Rev. Mod. Phys. **71**, S374–S382 (1999).

[30]D. J. Hornbaker, R. Albert, I. Albert, A.-L. Barabási, and P. Schiffer, "Why sand castles stand: An experimental study of wet granular media," Nature (London) **387**, 765 (1997).

[31]R. M. Nedderman, *Statics and Kinematics of Granular Materials* (Cambridge University Press, Cambridge, 1992).

[32]S. Flügge, *Practical Quantum Mechanics* (Springer, Berlin, 1994); S. Gasiorowicz, *Quantum Physics* (Wiley, New York, 1996).

[33]L. D. Landau and E. M. Lifshitz, *Quantum Mechanics, Nonrelativistic Theory*, Course of Theoretical Physics, Vol. 3 (Pergamon, New York, 1972).

[34]*Handbook of Mathematical Functions with Formulas, Graphs, and Tables*, Applied Mathematical Series Vol. 55, edited by M. Abramovitz and I. A. Stegun (National Bureau of Standards, Washington, D.C., 1964).

[35]A. A. Maradudin, "Surface acoustic waves," in *Nonequilibrium Phonon Dynamics*, edited by W. E. Bron (Plenum, New York, 1985), p. 395.

[36]R. J. Urick, *Principles of Underwater Sound* (McGraw Hill, New York, 1983).

[37]C. Chapman, *Fundamentals of Seismic Wave Propagation* (Cambridge University Press, Cambridge, 2004).

[38]B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990).

[39]H. M. Jager, S. R. Nagel, and R. P. Behringer, "The physics of granular materials," Phys. Today **49**, 32–38 (1996).

[40]J. H. Snoeijer, W. G. Ellenbroek, T. J. H. Vlugt, and M. Van Hecke, "Sheared force networks: Anisotropies, yielding, and geometry," Phys. Rev. Lett. **96**, 098001 (2006).

Ultrasonic oil-film thickness measurement: An angular spectrum approach to assess performance limits

Jie Zhang and Bruce W. Drinkwater[a)]
Department of Mechanical Engineering, University Walk, University of Bristol, Bristol BS8 1TR, United Kingdom

Rob S. Dwyer-Joyce
Department of Mechanical Engineering, Mappin Street, University of Sheffield, Sheffield S1 3JD, United Kingdom

The performance of ultrasonic oil-film thickness measurement in a ball bearing is quantified. A range of different viscosity oils (Shell T68, VG15, and VG5) are used to explore the lowest reflection coefficient and hence the thinnest oil-film thickness that the system can measure. The results show a minimum reflection coefficient of 0.07 for both oil VG15 and VG5 and 0.09 for oil T68 at 50 MHz. This corresponds to an oil-film thickness of 0.4 μm for T68 oil. An angular spectrum (or Fourier decomposition) approach is used to analyze the performance of this configuration. This models the interaction of component plane waves with the measurement system and quantifies the effect of the key parameters (transducer aperture, focal length, and center frequency). The simulation shows that for a focused transducer the reflection coefficient tends to a limiting value at small oil-film thickness. For the transducer used in this paper it is shown that the limiting reflection coefficient is 0.05 and the oil-film measurement errors increase as the reflection coefficient approaches this value. The implications for improved measurement systems are then discussed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713676]

I. INTRODUCTION

The function of lubricant oil in a machine element such as a bearing is to control friction and wear and hence provide smooth running and a satisfactory life. A large amount of recent lubrication research is devoted to the study, prevention, and monitoring of oil degradation.[1] Lubricant degradation in service can lead directly to machine element damage and machinery failure and so on-line monitoring is desirable. The oil-film thickness gives information about the operating condition of the oil and early warning of lubrication failure,[2] allowing life prediction or maintenance scheduling.

When an ultrasonic pulse strikes a very thin layer of lubricant (referred to in this paper as the oil-film) in a bearing system, the film behaves mechanically as a spring.[3] The reflected ultrasonic pulse is then a function of the oil-film stiffness, which in turn depends on the oil-film thickness and the elastic properties of the surrounding media. In this way a simple spring-layer model can be used to extract oil-film thickness from the measured reflection coefficient. Dwyer-Joyce *et al.*[3] and Zhang *et al.*[4] used this technique to monitor the oil-film thickness in machine elements such as journal, thrust-pad, and ball bearings. The results were shown to agree well with models of the bearing performance.[4] The thinnest oil-film thickness measured was about 0.4 μm (corresponding to a reflection coefficient, $R=0.09$) for Shell T68 engine oil in a 6016 ball bearing system.[4] There is significant interest in measuring even thinner oil-films as much lubrica-

tion occurs under boundary lubrication conditions which leads to oil-film thicknesses in the range 1–100 nm.[1] If very low reflection coefficients (which correspond to very thin oil-films) could be measured then this regime could be explored, potentially leading to a range of interesting measurement devices.

In this paper, the angular spectrum technique is used to analyze the ultrasonic measurement of oil-films. This approach uses a spatial Fourier transform to decompose an arbitrary ultrasonic field into its component plane waves.[5] These plane waves, which propagate at different angles, can then be analyzed separately and eventually recomposed into an ultrasonic field by an inverse angular spectrum. The use of the angular spectrum to model the propagation of acoustic fields and the output of transducers has been widely considered.[6–10] A number of authors have also used the angular spectrum approach in conjunction with multilayered system models. For example Atalar[11] used the angular spectrum method to analyze the performance of acoustic microscopes. He used various simple angular dependent reflection coefficient functions to represent the interaction of the acoustic microscope beam with a surface. Moidu *et al.*[12] used a similar approach to model the inspection of adhesively bonded joints using planar and focused transducers. The adhesive joint was modeled using a spring model of the interface and showed good agreement with a number of normal and oblique incidence experiments. Croce *et al.*[13] extended this analysis by using a full multilayered system model.

In this paper, the angular spectrum approach is used to model the output from highly focused ultrasonic transducers

---

[a)]Electronic mail: b.drinkwater@bristol.ac.uk
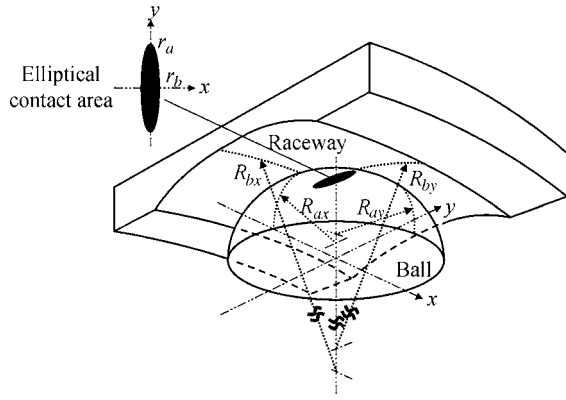
FIG. 1. Geometry of the contact between the outer-raceway of the bearing and ball.

and the interaction of the resultant acoustic fields with a thin oil-film. The primary aim is to quantify the performance of this measurement system and assess the effect of transducer parameters such as aperture, focal length, and center frequency. Experimentally, a minimum measurable oil-film reflection coefficient was observed and so the secondary aim of this paper is to explain this observation. Throughout the paper the discussion is focused on a specific bearing and transducer configuration, although it should be noted the approach presented is generally applicable.

II. BACKGROUND THEORY

A. Ball bearing lubrication

For a ball bearing, operating in the elastohydrodynamic lubrication regime, the oil-film thickness can be determined numerically from the regression equations of Dowson and Higginson.[14] They showed that the central film thickness, h_c, can be expressed as

$$\frac{h_c}{R'} = 2.69 \left(\frac{U\eta_0}{E'R'} \right)^{0.67} (\alpha E')^{0.53} \left(\frac{5W}{n_b E' R'^2} \right)^{-0.067}$$
$$\times (1 - 0.61 e^{-0.73m}), \quad (1)$$

where U is the mean surface speed, η_0 is the lubricant viscosity at the contact entry, α is the pressure-viscosity coefficient, W is the radial load on the whole bearing and n_b is the number of balls, m is the ellipticity parameter, E' is the reduced elastic modulus, and R' is the reduced radius of curvature given by

TABLE I. Parameters used to calculate the theoretical oil-film thickness in a 6016 ball bearing via Eq. (1) (Ref. 14).

Reduced modulus E' (GPa)	Reduced radius R' (m)	Pressure viscosity coefficient for T68 α (GPa^{-1})	Ellipticity parameter m	Simplified elliptical integrals n	Effective viscosity for T68 η_0 (N/m^2 s)
228	5.85E−3	20	11.5	1.014	0.2

TABLE II. Acoustic properties of lubricant oils and steel.

	Effective viscosity η_0 (N/m^2 s)	Density ρ (kg/m^3)	Longitudinal wave velocity c_l (m/s)	Bulk modulus B (GPa)
T68 at 0.1 MPa	0.2	876	1460	1.84
T68 at 1.5 GPa	0.2	1044	4500	21.2
VG15 at 0.1 MPa	0.07	940	1470	2.03
VG5 at 0.1 MPa	0.03	980	1480	2.14
Steel (EN24)	$\cdots$	7900	5900	172

$$\frac{1}{E'} = \frac{1}{2}\left[\frac{1 - \nu_a^2}{E_a} + \frac{1 - \nu_b^2}{E_b} \right], \quad \frac{1}{R'} = \frac{1}{R_{ax}} + \frac{1}{R_{bx}} + \frac{1}{R_{ay}} + \frac{1}{R_{by}}, \quad (2)$$

where E is Young's modulus and ν is Poisson's ratio. As shown in Fig. 1, the subscripts a and b refer to the two rolling elements (i.e., the ball and the raceway) and x and y refer to the coordinate axes. The contact area is elliptical in shape with the major (r_a) and minor (r_b) semicontact radii given by

$$r_a = \left(\frac{30 m^2 n W R'}{n_b \pi E'} \right)^{1/3}, \quad r_b = \left(\frac{30 n W R'}{n_b \pi m E'} \right)^{1/3}, \quad (3)$$

where n is a measure of the shape of the contact ellipse. In this paper a 6016 ball bearing was used and the parameters used in Eqs. (1)–(3) are given in Table I and the properties of the three oils used (T68, VG15, VG5) are shown in Table II. Note that the properties at 1.5 GPa are extremely difficult to measure and so are only accurately available for T68 oil.[15]

B. Reflection coefficient from an oil-film

Figure 2 shows a solid-lubricant layer-solid system, which represents the three-layer structure of a rolling element bearing. Following standard lubrication theory[1] it is assumed that the oil-film thickness is constant within the contact and equal to the central film thickness (i.e., $h=h_c$). Surfaces 1 and 2 represent the bearing raceway surface and ball surface, respectively. When an ultrasonic plane wave propagates through this structure, ultrasound will be reflected from both the top and bottom surfaces of the oil-film. However, as the oil-film is small compared to the wavelength it is modeled as a boundary condition between the raceway and ball.[16] The lubricant layer is then described by its normal and tangential stiffness, denoted by K_N and K_T, respectively. The normal stiffness of the lubricant layer can be simply related
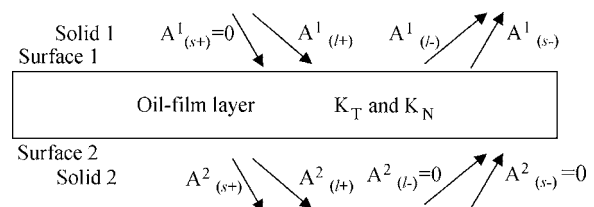
FIG. 2. Three-layer system model of the oil-film.

to its thickness, h, and bulk modulus, B (where $B = \rho_f c_f^2$, ρ_f is the density of the lubricant layer, and c_f is the velocity of the longitudinal wave in the lubricant layer) by[3]

$$K_N = \frac{B}{h}. \tag{4}$$

Figure 2 also shows the various waves which could exist where A is the amplitude of the plane wave, subscripts l and s refer to the longitudinal wave and shear wave, respectively, and superscripts 1 and 2 represent the media. Of particular interest to this paper are $A^1_{(l+)}$, the incident longitudinal wave in medium 1 and $A^1_{(l-)}$ the reflected longitudinal wave in medium 1. In this notation the longitudinal wave reflection coefficient in medium 1 is given by

$$R(\omega, \theta) = \frac{A^1_{(l-)}}{A^1_{(l+)}}. \tag{5}$$

Assuming that the half-spaces on either side of the lubricant layer have identical acoustic properties, the amplitudes of the various plane waves are related by,[16–18]

$$\begin{Bmatrix} A^2_{(l+)} \\ A^2_{(l-)} \\ A^2_{(s+)} \\ A^2_{(s-)} \end{Bmatrix} = [D]^{-1}[S][D] \begin{Bmatrix} A^1_{(l+)} \\ A^1_{(l-)} \\ A^1_{(s+)} \\ A^1_{(s-)} \end{Bmatrix}, \tag{6}$$

where $[D]$ defines the relationship between the wave amplitudes and the normal and shear stresses and displacements and $[S]$ describes the spring boundary condition between the two media. Matrices $[D]$ and $[S]$ are given by

$$[D] = \begin{bmatrix} c_l C_i C_{2\beta} & c_l C_i C_{2\beta} & -2sc_s^2 C_i C_2 & 2sc_s^2 C_i C_2 \\ 2sc_s^2 C_i C_1 & -2sc_s^2 C_i C_1 & c_s C_i C_{2\beta} & c_s C_i C_{2\beta} \\ c_l s & c_l s & C_2 & -C_2 \\ C_1 & -C_1 & -c_s s & -c_s s \end{bmatrix}, \tag{7}$$

$$[S] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \dfrac{1}{K_T} & 1 & 0 \\ \dfrac{1}{K_N} & 0 & 0 & 1 \end{bmatrix}, \tag{8}$$

where $C_1 = (1 - c_l^2 s^2)^{1/2}$, $C_2 = (1 - c_s^2 s^2)^{1/2}$, $C_{2\beta} = 1 - 2c_s^2 s^2$, $C_i = i\omega\rho$, $s = \sin(\theta_l)/c_l = \sin(\theta_s)/c_s$, ω is the center frequency of the plane wave, and θ is its incident angle, defined with respect to the surface normal. In this way the refraction and mode conversions of both longitudinal and shear waves are modeled and so Eqs. (6)–(8) could be applied to any incident or reflected wave mode if required. In this paper only longitudinal incident and reflected waves are considered as the transducer used was only sensitive to this wave type.
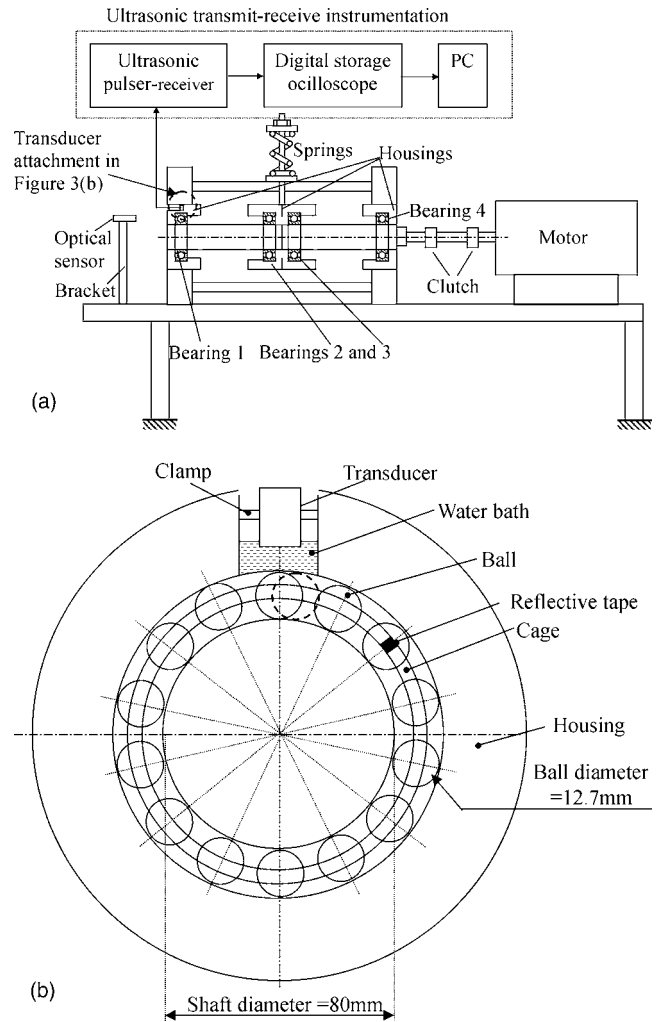
FIG. 3. (a) Schematic diagram of the experimental apparatus made up of four 6016 ball bearings and (b) transducer attachment and bearing geometry.

Note that for the case of a longitudinal wave, normally incident on a lubricant layer, Eqs. (4)–(8) can be simplified to calculate film thickness as,[3]

$$h = \frac{2B}{\omega Z} \sqrt{\frac{|R_n(\omega)|^2}{1 - |R_n(\omega)|^2}}, \tag{9}$$

where Z is the acoustic impedance of the media surrounding the lubricant film and $R_n(\omega)$ is the amplitude of the normal incident plane wave reflection coefficient. Equation (9) will suffer from inaccuracies as R_n approaches unity.

III. BALL BEARING EXPERIMENTAL APPARATUS

Figure 3(a) shows the experimental apparatus used to measure the ultrasonic reflection coefficient from an oil-film in a 6016 ball bearing system (supplied by NSK, UK). This bearing was selected as an example of a common medium sized rolling element (ball) bearing. A rotating shaft of 80 mm diameter was supported on four 6016 ball bearings lubricated via a total loss gravity feed system. Bearings 1 and 4 were fitted to the ends of the shaft and fixed into rigid housings. Vertically upwards radial loads were applied to the central region of the shaft through bearings 2 and 3 via an arrangement of springs. This meant that in bearings 1 and 4

TABLE III. Experimental and modeled transducer parameters.

	Center frequency f_c (MHz)	Focal length F (mm)	Active element diameter D (mm)	Taper window coefficient β
Experiment	50	23 (in water)	6.35	n/a
Model	50	4.5 (in steel)	5	0.9

the balls at the top of the raceway were the most heavily loaded. The rotary shaft speed was controllable in the range 100–2900 rpm by a 7.5 kW electric motor. This control of load and speed then enabled control of the resultant oil-film thickness via Eq. (1).

An optical sensor was used, both to allow accurate triggering of the ultrasonic instrumentation and to measure shaft speed. This was triggered by reflective tape attached to the ball cage (which rotates at half the shaft speed). Bearing 1 was instrumented with the ultrasonic measurement system that is shown in Fig. 3(b). A focused, longitudinal wave piezoelectric ultrasonic transducer was rigidly mounted in the housing such that it was normal to the top surface of the outer raceway. This transducer acted as both an emitter and receiver (pulse-echo mode), and its parameters shown in Table III. The transducer was focused on the outer raceway (4.5 mm thickness) and designed to achieve a focal zone size smaller than the width of the lubricated contact region. An ultrasonic pulser-receiver (Panametrics 5072PR) was used to excite the ultrasonic transducer, receive and amplify the reflected signals which were then passed to a digital oscilloscope (sampling frequency 5 GHz) and PC for storage and analysis. The temperature was monitored throughout the experiment and measurements taken only after a steady state was reached.

The reflective tape attached to the bearing cage is also shown in Fig. 3(b). When this tape passed the optical sensor it generated a 5 V positive trigger pulse. This pulse was used to trigger a signal generator (Agilent 33220A). After the addition of an adjustable delay the signal generator then triggered the pulser-receiver at its maximum pulse repetition frequency, which was 20 kHz. In this way, the lubricated contact was mapped spatially as the ball bearing passed under the measurement position. The minimum reflection coefficient was then selected for analysis and assumed to originate from the position of minimum oil-film thickness, i.e., the center of the lubricated contact.

The central reflection coefficient was measured by comparing the signal reflected from the oil-film, $A_m(\omega)$, with that from a reference interface, $A_{ref}(\omega)$,

$$R(\omega) = \frac{A_m(\omega)}{A_{ref}(\omega)} R_{ref}(\omega), \qquad (10)$$

where $R_{ref}(\omega)$ is the reflection coefficient of the reference interface, which in this case was a steel-air interface obtained before the lubricant was introduced. Care was taken to ensure that the alignment of the transducer with respect to the bearing was maintained between the steel-air and
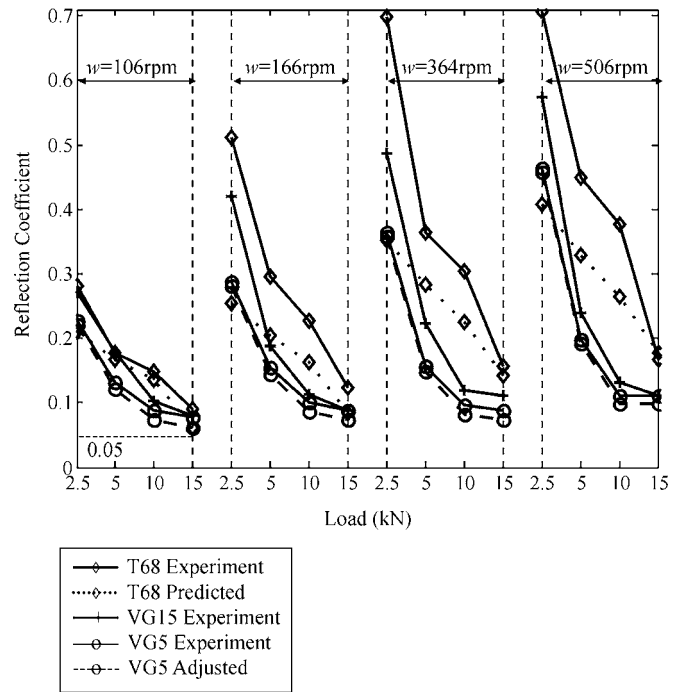
FIG. 4. Experimentally measured reflection coefficients for three oils (T68, VG15, and VG5) at various operating conditions. The dashed line is the reflection coefficient for oil VG5 after calibration and the dotted line is a prediction, based on Eq. (1), for the oil T68.

oil-film measurements. This reflection coefficient was then used in Eq. (9) to obtain the oil-film thickness.

Three oils with different viscosities (Shell T68, VG15, VG5) were used to explore the minimum measurable oil-film thickness. In general, oils with lower viscosities generate thinner oil-films for the same operating conditions (i.e., bearing load and speed). For this reason it was thought that the lowest reflection coefficient would result from the lowest viscosity oil.

IV. RESULTS

Figure 4 shows measurements of central reflection coefficient for the three oils (Shell T68, VG15, VG5) for a range of different operating speeds (w) and loads (W). From Fig. 4 it can be seen that the reflection coefficient decreases with increased load for all oils. It can also be seen that for a given oil the reflection coefficient decreases as speed decreases and that the lower viscosity oils exhibit lower reflection coefficients. Qualitatively, all these trends follow the changes in oil-film thickness predicted by Eq. (1), bearing in mind that a lower oil-film thickness is the cause of a lower reflection coefficient. However, this paper will contend that the reflection coefficient tends to a limiting value rather than tending to zero at high load and low speed as Eq. (1) would suggest.

V. DISCUSSION

The following considers the effect of the transducer parameters such as aperture, D, focal length, F, and center frequency, f_c, on oil-film thickness measurement by using
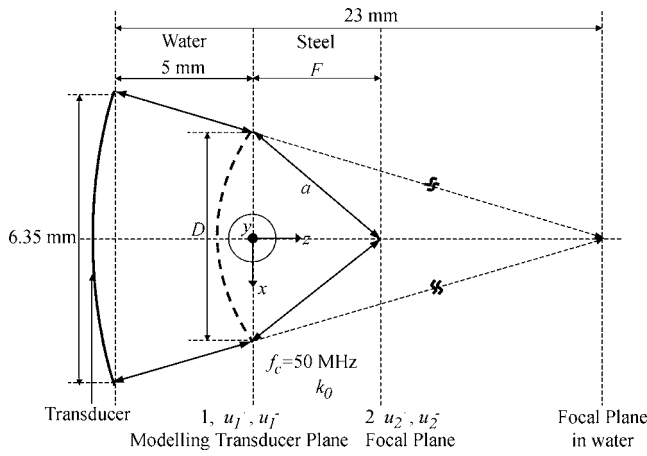
FIG. 5. Geometry of the focused transducer used for the angular spectrum simulation.

the angular spectrum approach.[6] The spring-layer model [i.e., Eqs. (5)–(8)] is used to model the wave interactions with the oil-film.

A. Angular spectrum approach

The geometry of the measurement system used for the modeling is shown in Fig. 5. For simplicity, the transducer is modeled as if it were directly coupled to the steel of the bearing shell. This modeled transducer is then forced to focus on the lubricated raceway-ball interface and so the need to model the propagation in the water is removed. The aperture size is adjusted by a simple trigonometric calculation to account for the small focusing effect in the water. In this way the 6.25-mm-diam transducer with a focal length of 23 mm in water used experimentally (see Table III) becomes a 5-mm-diam transducer of focal length in steel of 4.5 mm in direct contact with the steel (see Table III). In Fig. 5, the planes labeled 1 and 2 represent the modeled transducer plane and focal plane, respectively, and plane 2 is also the plane of the oil-film. In the discussion that follows, the superscripts + and − refer to fields propagating in the +z and −z directions, respectively.

The axial pressure of a focused circular transducer is given by Ref. 19 as

$$
p = p_0 \left| \frac{2}{1 - \dfrac{z}{a}} \right| \times \left| \sin \left[\frac{\pi}{\lambda} \left(\sqrt{\left(z - a + \sqrt{a^2 - \frac{D^2}{4}}\right)^2 + \frac{D^2}{4}} - z \right) \right] \right|,
$$
(11)

where a is the radius of curvature of the transducer, p_0 is the pressure when $z=a$, λ is the wavelength, D is the aperture of the transducer, and $a>D/2$ is assumed. If the focal length, F, defined as the distance at which pressure is a maximum then be this equation can then be used to calculate the effective radius of curvature that results in the desired focal length (i.e., 4.5 mm in this case and equal to the bearing raceway thickness in general).

The displacement field at plane 1 due to the transducer is then modeled by the following two-dimensional windowing function:

$$
u_1^+(x,y) = \begin{cases} 1 \\ 0.5 \times \left[\cos\left(\dfrac{\pi\sqrt{x^2+y^2}}{D/2} \right) + 1 \right] \\ 0 \end{cases},
$$

when

$$
\sqrt{x^2+y^2} < = \beta\frac{D}{2}, \quad \beta\frac{D}{2} < \sqrt{x^2+y^2} < \frac{D}{2},
$$
(12)

$$
\sqrt{x^2+y^2} > = \beta\frac{D}{2},
$$

where the coefficient β defines the extent of the piston-like region of the transducer (in this paper $\beta=0.9$). This taper was used to represent a real transducer as well as reducing the amplitude of the high spatial frequency components that result from the use of a pure piston source (which is essentially a rectangular windowing function). It should be noted that the resultant predicted transducer response was found to be relatively insensitive to the precise value of β used. Following the approach of Ref. 6 the angular spectrum at plane 1 $U_1^+(k_x,k_y)$ is obtained by taking a spatial Fourier transform (denoted by FT) of the displacement field,

$$
U_1^+(k_x,k_y) = \mathrm{FT}(u_1^+(x_1,y_1))
$$
$$
= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} u_1^+(x_1,y_1)
$$
$$
\times \exp(-j(k_x x_1 + k_y y_1))dx_1 dy_1,
$$
(13)

where k_x and k_y are the wave numbers in the x and y directions, respectively. In practice Eq. (13) is implemented using a fast Fourier transform routine on an array of data representing the discretized spatial distribution of displacements. Equation (13) is then multiplied by a phase term to account for propagation from plane 1 to plane 2 (i.e., the focal plane),

$$
U_2^+(k_x,k_y) = U_1^+(k_x,k_y)\exp(-j(\sqrt{k_0^2 - \sqrt{k_x^2+k_y^2}})F),
$$
(14)

where k_0 is the wave number in the steel and $k_0=\omega/c_{\mathrm{steel}}$. If it is assumed that all the reflected ultrasound comes from the lubricated contact region then the reflected field can be found from Refs. 6 and 12,

$$
U_2^-(k_x,k_y) = U_2^+(k_x,k_y)R(\omega,\theta),
$$
(15)

where, from Eq. (5), $R(\omega,\theta)$ is the oil-film reflection coefficient for a plane wave at an angle, θ, with respect to the normal of plane 2,

$$
\theta = \sin^{-1}\left(\frac{\sqrt{k_x^2+k_y^2}}{k_0} \right).
$$

A further identical propagation term can then be used to simulate the propagation in the −z direction back to plane 1 to give
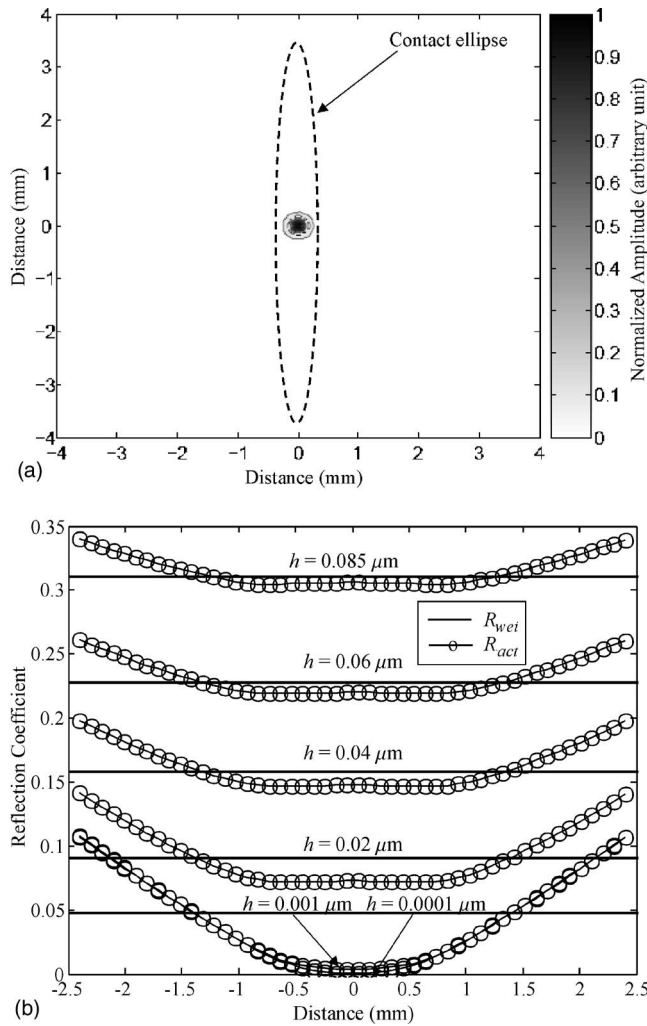
FIG. 6. Simulation results for the modeled transducer (defined in Table III). (a) Amplitude distribution at the focal plane. (b) Reflection coefficient distribution along the x axis for a number of oil-film thicknesses.

$$U_1^-(k_x,k_y) = U_1^+(k_x,k_y)$$

$$\times \exp(-j \cdot 2(\sqrt{k_0^2 - \sqrt{k_x^2 + k_y^2}})F)R(\omega,\theta). \quad (16)$$

The acoustic field reflected back to plane 1 and hence received by the transducer is then found by an inverse spatial FT of $U_1^-(k_x,k_y)$,

$$u_1^-(x,y) = \text{FT}^{-1}(U_1^-(k_x,k_y)) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U_1^-(k_x,k_y)$$

$$\times \exp(j(k_x x_1 + k_y y_1))dx_1 dy_1. \quad (17)$$

The amplitude recorded by the transducer, A, can then be calculated as the weighted sum of $u_1^-(x,y)$ over the transducer surface where the weighting function is that of the original transducer output, i.e., Eq. (12),

$$A = \sum_{x^2+y^2<(D/2)^2} u_1^-(x,y)u_1^+(x,y). \quad (18)$$

Figure 6 shows a simulation of the measurement system obtained using this approach. The acoustic properties of the media and transducer characteristics used in the simulation are shown in Tables II and III, respectively. A source plane

$(50 \times 50$ mm) was defined with a spatial sampling interval of 25 μm (i.e., $\Delta x = \Delta y = 25$ μm). Figure 6(a) shows the amplitude distribution at the focal plane (i.e., plane 2) obtained by Eq. (14) as well as the size of the lubricated contact region in a 6016 ball bearing for a 15 kN radial load from Eq. (3). From Fig. 6(a), it can be seen that the assumption that the majority of the acoustic field is within the contact ellipse is reasonable. Figure 6(b) shows the predicted reflection coefficient distribution and weighted sum reflection coefficient for various oil-films. To generate this figure Eqs. (17) and (18) were used to calculate the displacement distribution and weighted sum at the transducer. Simulations were performed for both the thin-film case and the reference case (i.e., steel-air) and used in Eq. (10) to compute reflection coefficient (using $R_{\text{ref}} = 1.0000$). For the reflection coefficient distribution, Eq. (10) was used for each spatial point. In this way the model replicated the experimental procedure. From Fig. 6(b) it can be seen that as the thickness of the oil-film increases, so the reflection coefficient distribution becomes flatter and begins to approach that of the weighted sum. It can also be seen from Fig. 6(b) that as the oil-film thickness decreases so the acoustic field distribution (and hence the sum) approach plateaux values. For example, oil-films of thickness 0.001 and 0.0001 μm can be seen to have barely distinguishable reflection coefficients. This suggests that, for the particular transducer modeled, there is a finite limiting reflection coefficient below which measurement is not possible. This in turn implies a limiting thickness as was observed experimentally in Fig. 4.

Simulations were performed for a range of oil-film thicknesses, and the weighted sum reflection coefficient is shown in Fig. 7(a), with a zoom of the important low thickness region shown in Fig. 7(b). Also shown in these figures is the normal incidence reflection coefficient (R_n) as a function of oil-film thickness. As the transducer is mounted at normal incidence it would be standard practice to use the normal incidence relationship to compute the oil-film thickness. For this reason the relative difference between the reflection coefficient measured by a focused transducer and the normal incidence plane wave reflection coefficient is defined as

$$\varepsilon = \left| \frac{h(R_{\text{wei}}) - h(R_n)}{h(R_n)} \right| \times 100 \% , \quad (19)$$

where $h(R)$ is the oil-film thickness obtained from Eq. (9), the normal incidence spring model. From Fig. 7(b) it can be seen that the smallest measurable reflection coefficient (labeled "R_{lim}") for the simulated transducer is 0.05. Figure 7(c) shows the relationship between the weighted sum reflection coefficient (R_{wei}) and the normal incident reflection coefficient (R_n) (labeled "calibration line"). As the calibration line departs from a line of unity gradient and zero intercept so the error between the normal incidence plane wave and the focused transducer reflection coefficient increases. For example, it can be seen from Fig. 7(c) that for a reflection coefficient of 0.1 measured with the focused transducer, the resultant oil-film thickness measurement error obtained from Eq. (9) is $\varepsilon = 17\%$. In principle, the calibration line shown in Fig. 7(c) could be used to compensate for this difference. As can be seen from

FIG. 7. Simulated reflection coefficient results from a range of oil-film thicknesses by using the modeled transducer (defined in Table III). (a) Reflection coefficients as a function of oil-film thickness. (b) Zoom in the left corner area in (a). (c) Calibration line.

TABLE IV. Reflection coefficient and oil-film thicknesses for various oils and various operating conditions. Also shown are oil-film thickness values from EHD theory (Ref. 14).

		T68	VG15	VG5
$w=106$ rpm, $W=15$ kN	R (Meas)	0.09	0.07	0.068
	h_c (μm)	0.41	0.31	0.3
	R (Calib)	0.07	0.05	0.046
	h_c (μm)	0.31	0.22	0.2
	EDH theory h_c (μm)	0.27	0.14	0.07
$w=506$ rpm, $W=15$ kN	R (Meas)	0.18	0.112	0.11
	h_c (μm)	0.81	0.5	0.49
	R (Calib)	0.177	0.099	0.097
	h_c (μm)	0.8	0.44	0.43
	EDH theory h_c (μm)	0.78	0.40	0.23

"effective" normal incidence reflection coefficient becomes 0.07, which corresponds to a 0.31 μm oil-film thickness. Note that in this case the EHD model value obtained from Eq. (1) is $h_c=0.27$ μm. Similar results are shown in Table IV for the other oils. The VG5 oil has the lowest viscosity of the three oils tested and so resulted in the thinnest oil-films (and therefore the lowest reflection coefficients). The above-described calibration approach has then been applied to the measured reflection coefficient data for this oil and the result is shown in Fig. 4. For example from Table IV it can be seen that at 15 kN and 106 rpm a reflection coefficient of 0.07 was measured which, after calibration, becomes 0.05. Although the acoustic properties of the VG5 oil under high pressures are not accurately known, if a similar relationship between pressure, density, and bulk modulus is assumed then this minimum reflection coefficient equates to $h_c=0.2$ μm. This is therefore the thinnest oil-film yet measured using ultrasound and this approach suggests a possible way forward to measure even thinner films. The transducer parameters responsible for the effects described in this section are now considered in more detail. In particular, experimental configurations which would enable the measurement of even thinner films are considered.

B. Variation of transducer parameters

For a transducer with $F=4.5$ mm and $f_c=50$ MHz, Fig. 8 shows the limiting reflection coefficient, $R_{\lim}$, and the reflection coefficients which result in various levels of oil-film thickness error, R_ε ($\varepsilon=2\%,5\%,10\%$), as a function of transducer aperture. From Fig. 8, it can be seen that both $R_{\lim}$ and R_ε increase with transducer aperture, D. This is because a focused transducer with a large aperture size can be thought of as emitting plane waves over a wide range of angles. Some of these plane waves are then obliquely incident on the oil-film and are hence sensitive to a different reflection coefficient than the normal incident plane waves. This implies that for oil-film thickness measurement, the

Table IV a reflection coefficient of 0.09 was measured for T68 oil at 15 kN and 106 rpm and this corresponds to a 0.41 μm oil-film thickness. If the calibration line in Fig. 7(c) is used to compensate for the focusing effect then the
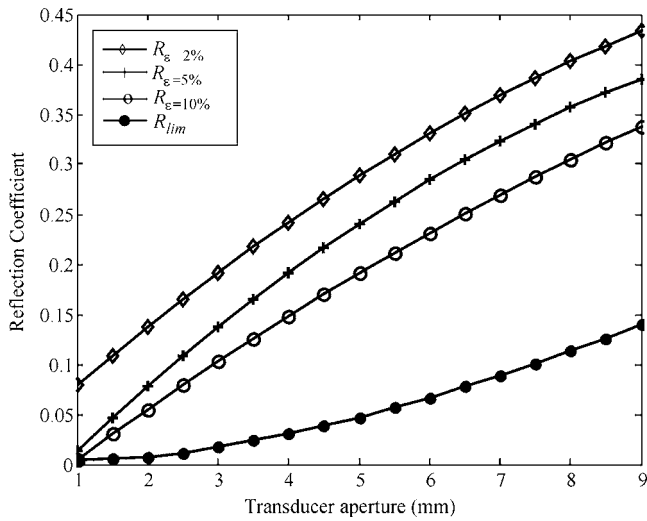
FIG. 8. Reflection coefficient as a function of transducer aperture, $F=4.5$ mm and $f_c=50$ MHz are assumed.

FIG. 10. Reflection coefficient as a function of center frequency of the transducer, $F=4.5$ mm and $D=5$ mm are assumed.

smaller aperture of the transducer, the better the measurement. However, the aperture must be such that the focal zone size is smaller than the lubricated contact.

For a given transducer aperture, $D=5$ mm and $f_c=50$ MHz, Fig. 9 again shows R_{lim} and R_ε for various error cases ($\varepsilon=2\%,5\%,10\%$), this time as a function of focal length. From Fig. 9, it can be seen that both R_{lim} and R_ε decrease as the focal length increases. As for the aperture the reason for this effect is that a focused transducer with a long focal length emits plane waves over a smaller range of angles. This implies that a large focal length is beneficial however, at 50 MHz the requirement for a small focal zone size forces the choice of a focal length to be as small as possible and hence equal to the bearing thickness.

For a given $F=4.5$ mm and $D=5$ mm, Fig. 10 shows R_{lim} and R_ε for various error cases ($\varepsilon=2\%,5\%,10\%$) as function of center frequency. From Fig. 10, it can be seen that R_{lim} is almost constant (0.05). However it should be noted that according to the spring-layer model [Eq. (9)], a

given reflection coefficient corresponds to a smaller oil-film thickness at a higher frequency. Hence, the smallest measurable oil-film thickness decreases with increased frequency.

From the analysis in this section, it can be seen that the ideal transducer for oil-film thickness measurement would have a high center frequency, a small aperture size, and long focal length. Figure 11 shows the performance of a possible transducer with these features i.e., $f_c=150$ MHz, $D=1$ mm, and $F=4.5$ mm focal length. From Fig. 11 it can be seen that it is possible to measure a reflection coefficient of 0.08 with 2% error. This corresponds to a 0.12 μm oil-film thickness for T68 oil in a 6016 bearing at 15 kN and 106 rpm.

VI. CONCLUSIONS

The use of a focused transducer to measure oil-film thickness has been shown to generate errors if not approached with care. These errors increase as oil-film thickness decreases, potentially limiting the ability of ultrasound to measure very thin oil-films. The angular spectrum ap-

FIG. 9. Reflection coefficient as a function of focal length of the transducer, $D=5$ mm and $f_c=50$ MHz are assumed.
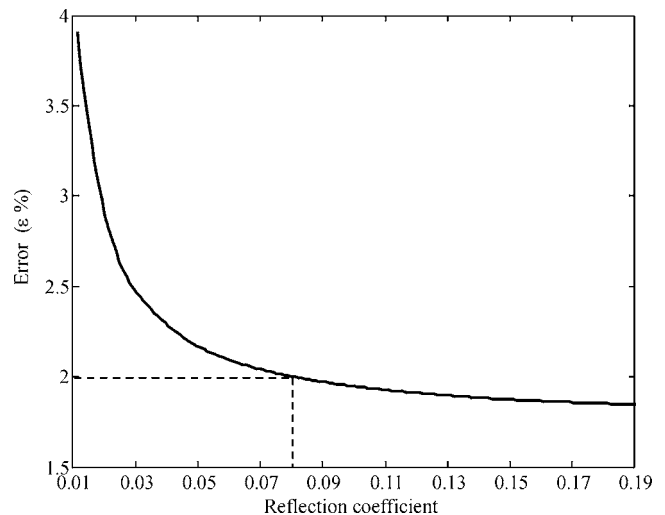
FIG. 11. Theoretical error from a focused transducer with 1 mm aperture, 150 MHz center frequency, and 4.5 mm focal length.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Zhang *et al.*: Ultrasonic oil-film thickness measurement   2619

proach has been used to model the measurement system and the interaction of the resultant plane waves with a spring layer. The analysis showed that oil-film thickness measurement errors decrease as center frequency and focal length increase and transducer aperture decreases. For a specific focused transducer (50 MHz center frequency, focal length of 4.5 mm, and aperture of 5 mm) the theoretical thickness measurement error was predicted to be 17% when the measured reflection coefficient was 0.1, and a finite reflection coefficient limit was reached at 0.05. This compared well with the experimentally observed reflection coefficient limit of 0.07. A calibration approach was described which allowed an oil-film thickness of 0.2 μm to be measured, thinner than had previously been possible. For the T68 oil which has well-known acoustic characteristics the experimental results were shown to be in improved agreement with known lubrication theory after calibration. The use of a small, high frequency transducer was then shown to offer improved performance on very thin oil-films.

ACKNOWLEDGMENT

[1] G. W. Stachowiak and A. W. Batchelor, *Engineering Tribology* (Butterworth Heinemann, Woburn, MA, 2001).

[2] J. Zhang, B. W. Drinkwater, and R. S. Dwyer-Joyce, "Monitoring of lubricant film failure in a ball bearing using ultrasound," Trans. ASME, J. Tribol. **128**, 622–618 (2006).

[3] R. S. Dwyer-Joyce, B. W. Drinkwater, and C. J. Donohoe, "The measurement of lubricant-film thickness using ultrasound," Proc. R. Soc. London, Ser. A **459**, 957–976 (2003).

[4] J. Zhang, B. W. Drinkwater, and R. S. Dwyer-Joyce, "The measurement of lubricant-film thickness in rolling element bearings using ultrasound," J. Acoust. Soc. Am. **119**, 863–871 (2006).

[5] J. W. Goodman, *Introduction to Fourier Optics* (McGraw Hill, New York, 1978).

[6] P. R. Stepanishen and K. C. Benjamin, "Forward and backward projection of acoustic fields using FFT methods," J. Acoust. Soc. Am. **71**, 803–812 (1982).

[7] P. T. Christopher and K. J. Parker, "New approaches to the linear propagation of acoustic fields," J. Acoust. Soc. Am. **90**, 507–521 (1991).

[8] D. P. Orofino and P. C. Pedersen, "Efficient angular spectrum decomposition of acoustic sources. I. Theory," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 238–249 (1993).

[9] C. J. Vecchio and P. A. Lewin, "Finite amplitude acoustic propagation modelling using the extended angular spectrum method," J. Acoust. Soc. Am. **94**, 2399–2408 (1994).

[10] P. Wu, R. Kazys, and T. Stepinski, "Optimal selection of parameters for the angular spectrum approach to numerically evaluate acoustic fields," J. Acoust. Soc. Am. **101**, 125–134 (1997).

[11] A. Atalar, "An angular-spectrum approach to contrast in reflection acoustic microscopy," J. Appl. Phys. **49**, 5130–5139 (1978).

[12] A. K. Moidu, A. N. Sinclair, and J. K. Spelt, "Nondestructive characterization of adhesive joint durability using ultrasonic reflection measurements," Res. Nondestruct. Eval. **11**, 81–95 (1999).

[13] R. Croce, P. Calmon, and L. Paradis, "Modeling of propagation and echo formation in a multilayered structure," Ultrasonics **38**, 537–541 (2000).

[14] D. Dowson and G. R. Higginson, *Elasto-Hydrodynamic Lubrication* (Pergamon, New York, 1977).

[15] O. B. Jacobson and A. P. Vinet, "Model for the influence of pressure on the bulk modulus and the influence of temperature on the solidification pressure for liquid lubricants," Trans. ASME, J. Tribol. **109**, 709–714 (1987).

[16] T. P. Pialucha and P. Cawley, "The detection of thin embedded layers using normal incidence ultrasound," Ultrasonics **32**, 431–440 (1994).

[17] M. J. S. Lowe, "Matrix techniques for modelling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).

[18] S. I. Rokhlin and Y. J. Wang, "Analysis of boundary conditions for elastic wave interaction with an interface between two solids," J. Acoust. Soc. Am. **89**, 503–515 (1990).

[19] H. T. O'Neil, "Theory of focusing radiators," J. Acoust. Soc. Am. **21**, 516–526 (1949).

Measurement of focused ultrasonic fields using a scanning laser vibrometer

Yuebing Wang
Hangzhou Applied Acoustics Research Institute, Fuyang, Hangzhou, 311400, China

John Tyrer
Loughborough University, Loughborough Leicestershire LE11 3TU, United Kingdom

Ping Zhihong and Wang Shiquan
Hangzhou Applied Acoustics Research Institute, Fuyang, Hangzhou, 311400, China

With the development of optical techniques, scanning laser vibrometers have been applied successfully in measuring particle velocities and distributions in ultrasonic fields. In this paper, to develop the optical interferometry in measuring focused fields with small amplitude, the "effective" refractive index used for plane waves and extended for spherical waves is presented, the piezo-optic effect as a function of the incident angle of the laser beam is simulated, and the ultrasonic field produced by a concave spherical transducer is calculated numerically around its focal region. To verify the feasibility of the optical method in detecting focused ultrasonic fields, a measurement system was set up that utilized both a scanning laser vibrometer and a membrane hydrophone. Measurements were made in different zones of a focusing transducer, and good results were acquired from the optical interferometry in regions where acoustic waves travel in plane form or spherical form. The data obtained from the optical method are used to reconstruct acoustic fields, and it is found that the focal plane, the maximum pressure, and the beamwidth of the transducer can be forecasted accurately. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713708]

I. INTRODUCTION

Focusing transducers have been widely used as medical ultrasonic devices. To evaluate acoustic properties of these transducers, various needle hydrophones have been designed.[1,2] However, a hydrophone has a limited spatial resolution, which is determined by the size of its active element. To avoid the error due to spatial averaging, the effective radius of a needle hydrophone should be smaller than one quarter of the acoustic wavelength or one third of the −6-dB beamwidth of the focused acoustic field.[3,4] On the other hand, to obtain high-resolution images or improve therapeutic effect, focusing transducers are made with focal regions less than 1 mm, hence, traditional methods are facing challenges in measuring high focused ultrasonic fields.

Optical methods have long been applied in detection of ultrasound and there are a few techniques that can provide output in response to ultrasonic fields.[5–9] As a noninvasive method, the optical detection has the advantages of measuring ultrasound fields in their unperturbed states and avoiding the damage to measurement devices. Using a thin membrane (pellicle), the optical interferometry can be adopted to measure particle velocities in an ultrasonic field.[10–12] Nowadays, commercial laser Doppler vibrometers with excellent performance have been developed that can achieve much higher temporal and spatial resolution than by the hydrophone method, moreover, a laser scanning vibrometer can make acoustic field measurement more convenient.[13,14]

When the pellicle-based measurement method is adopted, the laser beam of a vibrometer will interact with acoustic waves passing through the pellicle. As the vibrometer interprets perturbations on the laser beam as velocities, the piezo-optic effect of the water medium must be taken into account so that the particle velocity on the pellicle can be identified. In a plane wave field, when the laser beam of a vibrometer is in parallel with the acoustic beam, an "effective" refractive index can be introduced to make corrections for the piezo-optic effect.[15] However, in other forms of acoustic fields, or when a vibrometer is applied to make a scanning measurement, the piezo-optic effect will cause a complicated modulation on the laser beam penetrating into acoustic fields. To take the optical interferometry as an effective technique in measuring focused fields with small amplitude, the piezo-optic effect should be further studied and results from the optical method carefully investigated.

In the theoretical part of the paper, the effective refractive index for a plane wave is reviewed, and further considerations are given for the piezo-optic interaction of spherical waves and for the case where a laser beam is off the acoustic axis of a projector. Finally, the pressure distributions in amplitude and phase of a typical focusing transducer are simulated. In the experimental part, the methodology and facility for measuring focused fields are first described, and the results from both the optical vibrometer and the membrane hydrophone are compared on different zones of a focusing radiator. Finally, the pressure distribution on the focal plane is forecasted with the data from a scanning vibrometer and compared with measurement results obtained by a needle hydrophone. These analyses and measurements lead to a few conclusions about the application of the optical interferom-
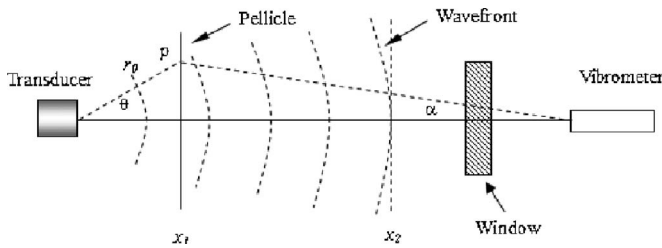
FIG. 1. Pellicle-based optical interferometry for ultrasonic field measurement.

etry in measuring focused fields.

II. THEORETICAL CONSIDERATION

A. Refractive index n_0 and piezo-optic coefficient n_1

The piezo-optic coefficient is an inherent property of a medium defined as $n_1 = (\partial n / \partial p)_s$, which describes the variation of the refractive index with its ambient pressure under adiabatic circumstance. Refractive index n_0 and piezo-optic coefficient n_1 of pure water have already been investigated in many experiments, usually, $n_0 = 1.33$ and $n_1 = 0.32$ are adopted under the conditions of normal atmospheric pressure, water temperature $T = 20\ ^\circ C$, and laser wavelength $\lambda = 632$ nm.[16–19]

When a vibrometer is used to detect velocities on a pellicle, its laser beam will propagate in the acoustic field and interact with acoustic waves passing through the pellicle. As the vibrometer interprets local perturbations as velocities, the "piezo-optic coefficient" of the water medium must be derived to determine the velocity on the pellicle.

B. Calculation of particle velocity

Assume the geometrical arrangement for the optical detection is depicted in Fig. 1, in which the unsteered laser beam from a vibrometer is coincident with the acoustic axis of a transducer. A pellicle is placed at the plane x_1, and the acoustic wavefront travels to the position x_2 at time t. Under the condition that the amplitude of the acoustic wave is much smaller compared with the acoustic wavelength, the optical path length $q(t)$ detected by the vibrometer at time t can be denoted by[20–22]

$$q(t) = -2n_0a(t) + \frac{2n_1}{\rho c^2} \int_{x_1}^{x_2} p(x,t)dx, \quad (1)$$

where $a(t)$ is the amplitude of the acoustic wave at time t, $p(x,t)$ is the acoustic pressure at point x and time t, ρ is the density of water, and c is the speed of sound in water.

For linear plane waves, the velocity v detected by the vibrometer can be deduced from Eq. (1), as

$$v(t) = -2(n_0 - n_1)u(x_1,t), \quad (2)$$

where $u(x_1 t)$ is the velocity on the pellicle at time t.

Equation (2) shows that the measured velocity v is related to the velocity of the pellicle u by a constant $n_0 - n_1$, thus $n_p^* = n_0 - n_1$ is defined as the effective refractive index of plane waves.

However, the plane wave can only be achieved under ideal conditions, usually acoustic waves radiated from transducers propagate in spherical form. In Fig. 1, supposing the acoustic center of a transducer is placed on the origin of the x axis, the laser beam from a vibrometer is coincident with the x axis and is incident perpendicularly on the pellicle at x_1, then Eq. (1) can be written as

$$q(t) = -2n_0a(t) + \frac{2n_1}{\rho c^2}A_0 \exp i\omega t \int_{x_1}^{x_2} \frac{\exp - ikx}{x}\, dx, \quad (3)$$

where A_0 is the amplitude of acoustic pressure, ω is the angular frequency, and k is the wave number.

From Eq. (3), the ratio of the measured velocity v to the pellicle velocity u can be written as

$$n_s^* = n_0 - n_1 \frac{(kx_1)^2}{kx_1 - i} \exp ikx_1$$

$$\times \left(\frac{\exp - ikx_2}{kx_2} + i \int_{x_1}^{x_2} \frac{\exp - ikx}{x}\, dx \right) \quad (4)$$

Though the expression for n_s^* is complex, numerical calculation shows that Eq. (4) can be simplified as $n_s^* = n_0 - n_1$ when the pellicle is put three wavelengths away from the transducer ($x_1 > 3\lambda$). Hence, n_s^* can be regarded as the effective refractive index of spherical waves.

Furthermore, if the laser beam of the vibrometer scans an angle α from the x axis and falls aslant at point p on the pellicle (the angle between the x axis and the acoustic beam will be θ shown in Fig. 1), the effective refractive index becomes

$$n_s^* = n_0 \frac{\cos \theta}{\cos \alpha} - n_1 \frac{(kr_0)^2}{kr_0 - i} \exp ikr_0 \left[\frac{\exp - ikx_2}{kx_2\sqrt{1 - (d/x_2)^2}} \right.$$

$$\left. + i \int_{x_1/\cos(\theta)}^{x_2} \frac{\exp - ikr}{r} \frac{\exp - ikx_2}{\sqrt{1 - (d/r)^2}}\, dr \right], \quad (5)$$

where $d = x_1 \cos \alpha(\tan \theta + \tan \alpha)$ and r_0 is the distance from the projector to point p.

Equation (5) shows that n_s^* is decided by many factors; to reveal the relationship between n_s^* and the angle α, assuming $x_1 = 0.20$ m, $x_2 = 0.40$ m, and $f = 520$ kHz, the curve of n_s^* as a function of angle α is given in Fig. 2, which shows that the effective refractive index n_s^* would drop obviously from its peak value 1.01 with the increase of the scanning angle. To reduce the influence of piezo-optic effect on the output signals of a vibrometer, scanning measurements should be carried out within small angles.

C. Calculation of acoustic pressure

For linear plane waves, the acoustic pressure p can be derived from the detected velocity v by

$$p(t) = -\frac{\rho c v(t)}{2n_p^*}. \quad (6)$$

For small amplitude spherical waves, the acoustic pressure p is given by[23]
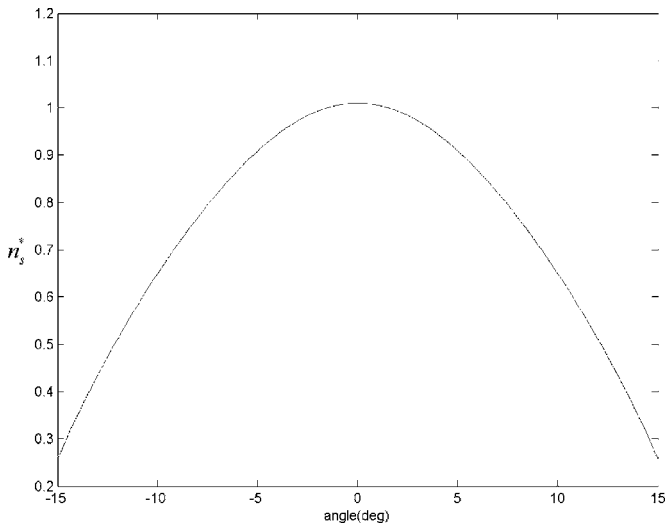
FIG. 2. n_s^* as the function of scan angle α ($l_0 = 0.20$ m, $l_1 = 0.40$ m, and $f = 520$ kHz).

$$p(x,t) = \frac{\rho c u(x,t)}{(1 + i\lambda/2\pi x)}. \qquad (7)$$

Hence, the acoustic pressure p can be expressed with the detected velocity v

$$p(x,t) = -\frac{\rho c v(t)}{2n_s^*(1 + i\lambda/2\pi x)}. \qquad (8)$$

When $x > 3\lambda$, the acoustic pressure p can be calculated with Eq. (6) and errors will be negligible.

D. Acoustic field of focusing transducer

Assuming that a concave spherical transducer has a uniform normal velocity v_0, and the radius a of its circular boundary is large relative to the acoustic wavelength and the depth of the concave surface, approximate expressions can be derived from calculating its pressure distribution.[24,25]

If the x axis of cylindrical coordinates is coincident with the acoustic axis of the concave spherical transducer, with the origin at its inner surface, and r represents the distance of a field point off the x axis, the velocity at the field point (x,r), can be given as[26]

$$u(x,r) = \frac{v_0}{x} \exp\left(\frac{-ikr^2}{2x}\right) \int_{r'=0}^{a} \exp\left[\frac{-ikr'^2}{2}\left(\frac{1}{x}\right.\right.$$

$$\left.\left. -\frac{1}{R}\right)\right] J_0\left(\frac{krr'}{x}\right) r' \, dr', \qquad (9)$$

where R is the curvature radius of the spherical shell and J_0 and is the Bessel function of zeroth order.

Based on Eq. (9), the acoustic field of a concave spherical transducer is calculated numerically and its amplitude and phase distributions on an x-r plane are given in Fig. 3, which shows that acoustic waves approach its focus in plane form and then spread as spherical form in the far Fresnel zone. These regions can be used as "detection windows" for the optical interferometry.

(a)

(b)
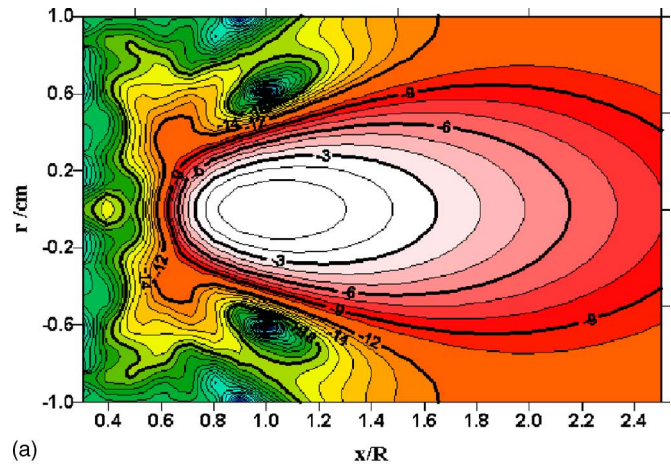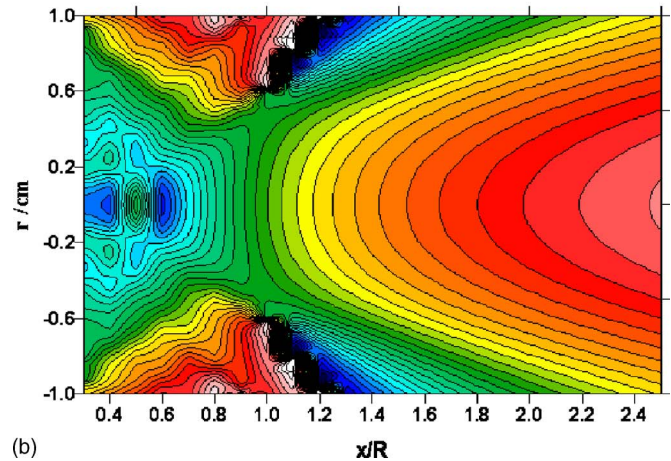
FIG. 3. (Color online) Distributions of amplitude (a) and phase (b) of a concave spherical transducer on an x-r plane ($a = 45$ mm, $R = 60$ cm, $f = 1$ MHz).

III. EXPERIMENT METHODOLOGY

To investigate the optical method for detection of focused fields, an experimental system was set up, as shown in Fig. 4.

In the system, a Textronix AFG 2020 generator was used to send out tone-burst signals, which were amplified by an ENI 2100L power amplifier and then input to a focusing transducer with a resonant frequency of 520 kHz. The transducer was fixed to the wall of a water tank and its acoustic energy radiated into the water medium directly. A bilaminar shielded membrane hydrophone with a 1.0-mm-diam active
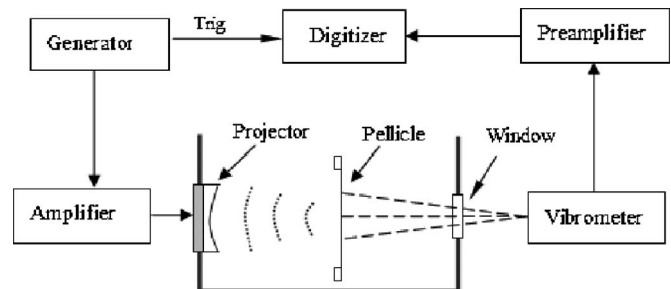
FIG. 4. Experimental system schematic diagram for focused field measurement.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Wang *et al.*: Optical measurement of focused ultrasonic fields   2623
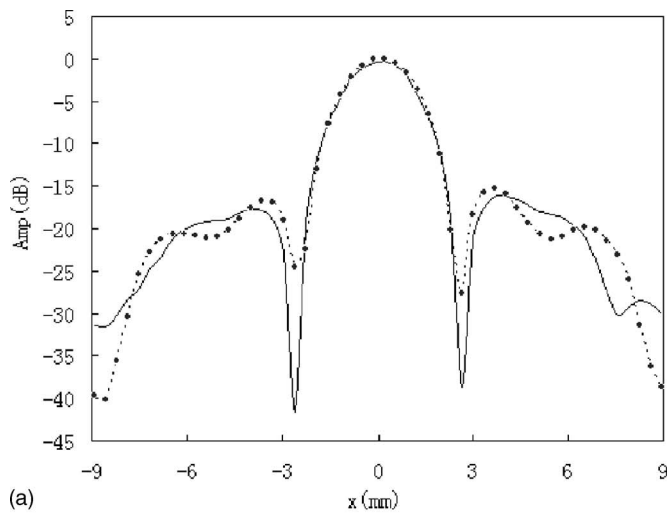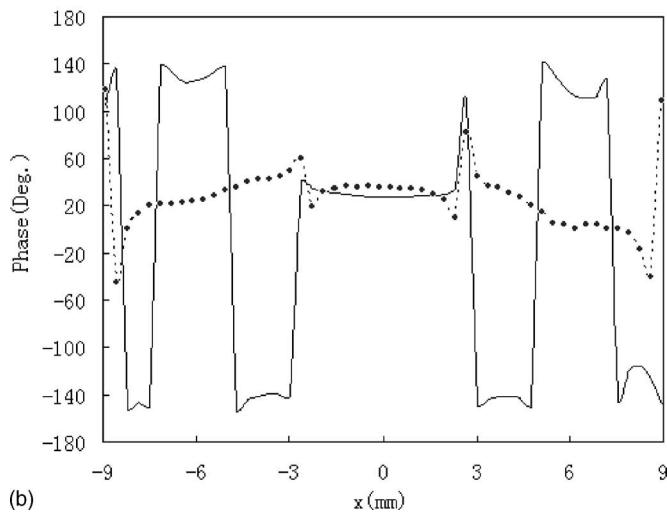
FIG. 5. Distributions on the focal plane in amplitude (a) and phase (b) from the vibrometer and the hydrophone:————hydrophone, -----vibrometer.

FIG. 6. Distributions on 30-mm plane in amplitude (a) and phase (b) from the vibrometer and the hydrophone:————hydrophone, -----vibrometer.

element was placed in the tank via a positioning system. Since the hydrophone consists of two layers of 25-μm-thick film stretched over an annular ring (100 mm in internal diameter) and is coated on either side in a gold layer, it can be used for reflecting laser beams as well. A Polytec scanning vibrometer (PSV-200) was placed on the wall opposite to the transducer and its unsteered laser beam was coincident with the beam axis of the transducer. This type of vibrometer comprises a scanning head (OFV-056) and a controller (OFV-3001). A displacement decoder OVD30 is installed in the controller that covers the frequency range from 50 kHz to 20 MHz and detects displacement from less than 1 Å up to 75 nm. The head of the vibrometer was placed 60 cm away from the pellicle; at this distance, the laser beam could be focused on the membrane into a 150-μm-diam spot, and, in scanning measurement, the largest deflection angle of the laser beam was less than 2.0 deg. Signals from the vibrometer and the hydrophone passing through the same preamplifier NF 5305 were analyzed using the digital scope Agilent 54833D.

In the experiment, the laser beam of the vibrometer was set to scan on the membrane hydrophone along a horizontal

line, and the acoustic pressure on the line was derived with Eq. (6). After the optical scanning ended, the membrane hydrophone was driven by the positioning system and the acoustic pressure on the same scanning line was measured, so the pressure distribution on the line could be compared to the optical method.

IV. MEASUREMENT RESULTS

The measurement is made on the focal plane 22 mm away from the front surface of the transducer; the results obtained by the two methods are shown in Fig. 5 in which the acoustic pressure is expressed in decibels and the maximum pressure 150 kPa is defined as 0 dB. The vibrometer and the hydrophone have different phase responses, which have not been calibrated yet; to investigate phase distributions, an extra 80 deg is added to the results from the vibrometer, so the phase values obtained by the two methods are nearly the same around the beam centerpoint.

Figure 5 shows the acoustic pressure obtained by the two methods is in good agreement inside the major lobe where acoustic waves travel in plane form, and in the −20-dB beamwidth region their difference is less than 5%.
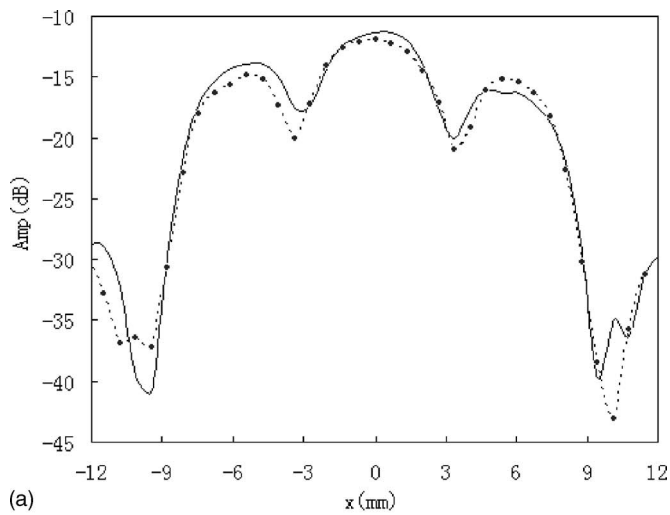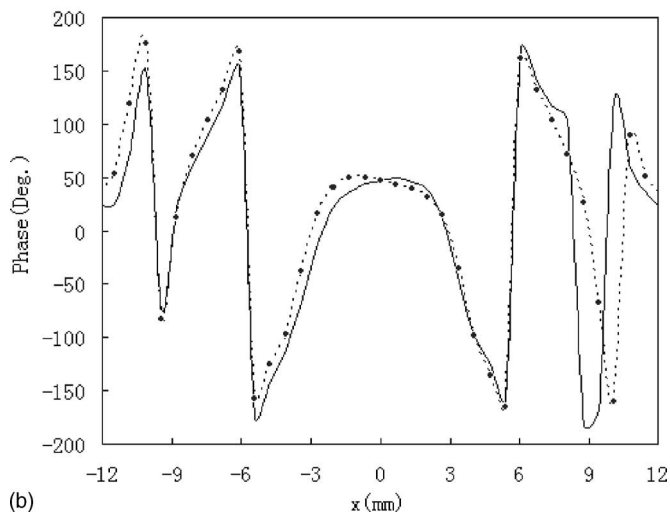
FIG. 7. Distributions on 35-mm plane in amplitude (a) and phase (b) from the vibrometer and the hydrophone:————hydrophone: ---- vibrometer.

FIG. 8. Distributions on 40-mm plane in amplitude (a) and phase (b) from the vibrometer and the hydrophone:————hydrophone, ----- vibrometer.

Outside the major lobe, quite different distributions would be observed in both phase and amplitude when phase shifting occurs.

When the membrane hydrophone was moved on the plane 30 mm away from the transducer's surface, acoustic pressure on the scanning line was measured and processed in the same way as described above, and the final results are given in Fig. 6.

On this plane, acoustic waves are in the process of changing their wavefronts; different from plane waves and spherical waves, the piezo-optic effect of the water medium takes place in a complicated manner, so acoustic pressure can not be calculated correctly by Eq. (6) and (8).

The membrane hydrophone was placed on positions 35 and 40 mm away from the transducer's surface. Pressure distributions were then measured and the processed results are given in Figs. 7 and 8. These figures show acoustic waves will spread in a quasi-spherical manner after they pass through the focal region, hence, acoustic pressure can be detected using the optical method. However, measurement errors larger than 1.0 dB will be produced due to the deviation of spherical waves in some regions.

V. RECONSTRUCTION OF ACOUSTIC FIELDS

The experiment confirms that an optical vibrometer can be used to detect focused fields in regions where acoustic waves propagate in spherical form, so it would be feasible to use the results from an optical scanning vibrometer to reconstruct the acoustic field on the focal plane.

On the plane 35 mm away from the transducer, the acoustic energy in the major lobe is found to spread into a 20-mm circular area in which approximatly 84% of the total power radiated is contained,[27] so a 25-mm square on the membrane hydrophone is chosen as the scanning area. With a spatial increment of 0.5 mm and sampling on a 51×51 square grid, acoustic pressure in amplitude and phase on 2601 points are measured as shown in Fig. 9. These data are used for forecasting acoustic fields.

The acoustic field around the focal region has been simulated. It is found that in the plane 22 mm away from the transducer, acoustic waves propagate in plane form and the amplitude reaches the maximum value of 141 kPa, which is 0.4 dB lower than the result obtained directly on the focal plane by the optical method. The reconstructed pressure distribution is shown in Fig. 10. The beamwidths at −3, −6, and

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Wang *et al.*: Optical measurement of focused ultrasonic fields   2625
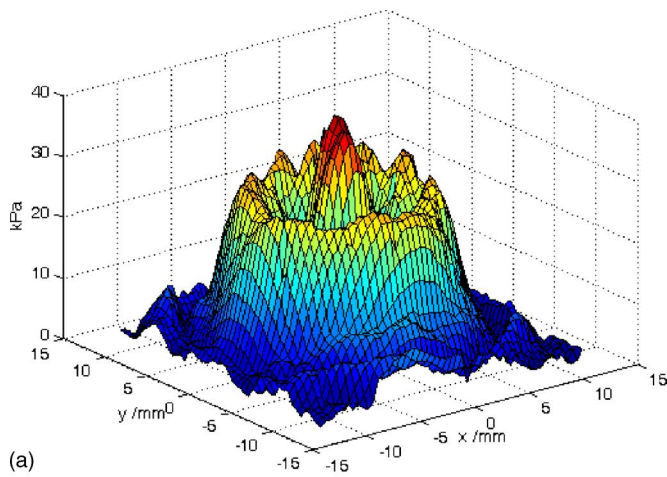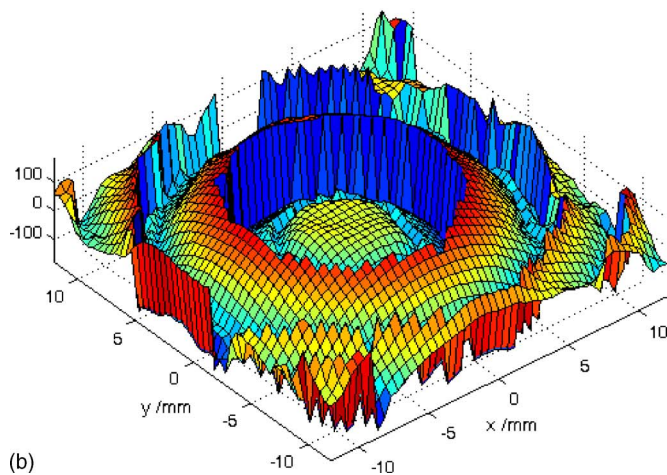
FIG. 9. (Color online) Distributions of amplitude (a) and phase (b) measured on the plane 35-mm away from the focusing transducer using a scanning vibrometer.

−12 dB (denoted as w_3, w_6, and w_{12}) are 2.0, 2.7, and 3.7 mm, respectively, while by the optical method the w_3, w_6, and w_{12} are 2.1, 2.9, and 3.9 mm, respectively, as shown in Fig. 5.
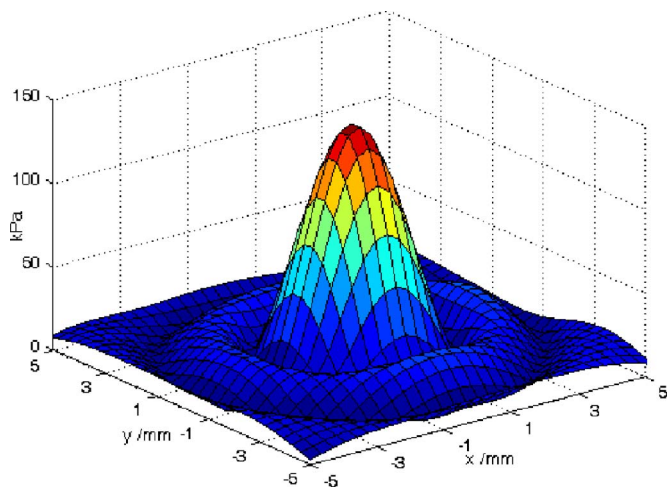
FIG. 10. (Color online) Reconstructed pressure distribution of the focal plane based on the data from a vibrometer and measured on the 35-mm plane (w_3=2.0 mm, w_6=2.7 mm, and w_{12}=3.7 mm).
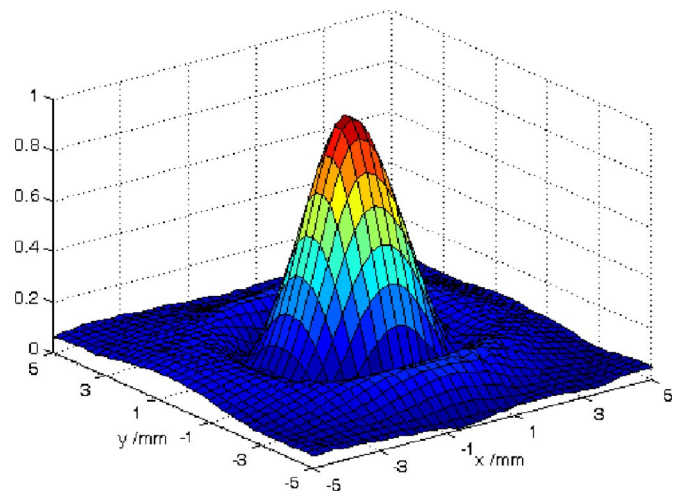
FIG. 11. (Color online) Normalized pressure distribution on the focal plane obtained by a needle hydrophone (w_3=2.0 mm, w_6=2.6 mm, and w_{12} =3.7 mm).

Finally, acoustic pressure on the focal plane is measured using a needle hydrophone and normalized results are shown in Fig. 11. The comparison between Figs. 10 and 11 confirms that the acoustic field inside the major lobe can be forecasted. However, the reconstruction is based on the data measured in a limited scanning area, the truncation influences the distribution inside major lobe region obviously, and an undesired bigger "side lobe" appears as shown in Fig. 10.

VI. CONCLUSION

When an optical vibrometer is utilized to measure ultrasonic fields, its output signal will be modulated by the piezo-optic effect of the medium. When used to measure a plane wave field, the influence of the piezo-optic effect can be removed with a constant n_p^* called the effective refractive index. Though the piezo-optic effect tends to be complicated for spherical waves, particle velocities in acoustic fields can be calculated with the same effective refractive index n_p^* when measurements are undertaken at places three wavelengths away from the source. In the scanning measurement, with the laser beam off the beam axis of a transducer, the effective refractive index drops off from its normal value, so the scanning measurement completed within small angles would be favorable for simplifying the analyses.

The theoretical simulation reveals that though a focusing transducer produces a complicated distribution in the near Fresnel zone, its acoustic waves pass the focus in the plane wave and then spread in spherical form in the far Fresnel zone. This behavior makes the optical method feasible in detecting focused acoustic fields.

The experiments have been conducted with a 520-kHz focusing transducer. On the focal plane, excellent agreement between the two methods is found in the −20-dB beamwidth region where acoustic waves travel in plane form. Outside the plane wave region, when acoustic waves begin to propagate with wavefronts varied, the optical method cannot provide good results. Over the focal region, acoustic waves

spread in quasi-spherical form; the optical method is feasible in most cases though a difference larger than 1.0 dB is observed.

The acoustic field on the focal plane is reconstructed using the data measured on a scanning area where acoustic waves spread in spherical form. When the scanning area covers the acoustic energy contained in the major lobe region on the focal plane, the position of the focal plane could be forecasted accurately, the maximum pressure is 0.4 dB lower, and the −6-dB beam width is 7% smaller than the results measured directly on the focal plane. However, in the side lobe region, acoustic pressure would be a little bigger due to the truncation of measurement region.

[1] P. A. Lewin, "Miniature piezoelectric polymer ultrasonic hydrophone probes," Ultrasonics **19**, 213–216 (1981).

[2] P. A. Lewin and R. C. Chivers, "Two miniature ceramic ultrasonic probes," J. Phys. E **14**, 1420–1424 (1981).

[3] R. C. Preston, *Output Measurements for Medical Ultrasound* (Springer-Verlag, New York, 1991).

[4] IEC 61102, "Measurement and characterization of ultrasonic fields using hydrophones in the frequency range 0.5 MHz to 15 MHz," (International Electrotechnical Commission, Geneva, 1991).

[5] R. Reibold and W. Molkenstruck, "Light diffraction tomography applied to the investigation of ultrasonic fields. Part 1: continuous waves," Acustica **56**, 180 (1984).

[6] J. P. Monchalin, "Optical detection of ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**(5), 485–499 (1986).

[7] P. A. Chinnery, V. F. Humphrey, and C. Beckett, "The Schlieren image of two-dimensional ultrasonic fields and cavity resonances," J. Acoust. Soc. Am. **101**(1), 250–256 (1997).

[8] T. A. Pitts and J. F. Greenleaf, "Three-dimensional optical measurement of instantaneous pressure," J. Acoust. Soc. Am. **108**(6), 2873–2883 (2000).

[9] A. R. Harland, J. N. Petzing, J. R. Tyrer, C. J. Bickley, S. P. Robinson, and R. C. Preston, "Application and assessment of laser Doppler velocimetry for underwater acoustic measurements," J. Sound Vib. **265**, 627–645 (2003).

[10] P. D. Theobald, S. P. Robinson, A. D. Thompson, R. C. Preston, P. A. Lepper, and W. Yuebing, "Technique for the calibration of hydrophones in the frequency range 10 to 600 kHz using a heterodyne interferometer and an acoustically compliant membrane," J. Acoust. Soc. Am. **118**(5), 3110–3116 (2005).

[11] D. Royer, N. Dubois, and M. Fink, "Optical probing of pulsed, focused ultrasonic fields using a heterodyne interferometer," Appl. Phys. Lett. **61**(2), 151–155 (1992).

[12] F. P. Higgins, S. J. Norton, and M. Linzer, "Optical interferometry visualization and computerized reconstruction of ultrasonic fields," J. Acoust. Soc. Am. **68**(4), 1169–1176 (1980).

[13] A. R. Harland, J. N. Petzing, and J. R. Tyrer, "Nonperturbing measurements of spatially distributed underwater acoustic fields using a scanning laser Doppler vibrometer," J. Acoust. Soc. Am. **115**, 187–195 (2004).

[14] Y. Wang, P. Theobald, J. Tyrer, and P. Lepper, "The application of scanning vibrometer in mapping ultrasound fields," J. Phys.: Conf. Ser. **1**, 167–173 (2004).

[15] D. R. Bacon, "Primary calibration of ultrasonic hydrophones using optical interferometry," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**(2), 152–161 (1988).

[16] W. A. Riley and W. R. Klein, "Piezo-optic coefficients of liquids," J. Acoust. Soc. Am. **42**, 1258–1261 (1967).

[17] H. Eisenberg, "Equation for the refractive index of water," J. Chem. Phys. **43**, 3887–3892 (1965).

[18] E. Reisler and H. Eisenberg, "Refractive indices and piezo-optic coefficients of deuterium oxide, methanol and pure water," J. Chem. Phys. **43**, 3875–3880 (1965).

[19] R. M. Waxler and C. E. Weir, "Effect of temperature and pressure on the refractive indices of benzene carbon tetrachloride and water," J. Res. Natl. Bur. Stand., Sect. A **67A**, 163–171 (1963).

[20] D. R. Bacon, "The improvement and evaluation of a laser interferometer for the absolute measurement of ultrasonic displacements in the frequency range up to 15 MHz," NPL Acoustics Report AC109, November 1986 (National Physical Lab., Middlesex, UK).

[21] C. B. Scruby and L. E. Drain, *Laser Ultrasonics* (Hilger, Bristol, 1990).

[22] D. R. Bacon, R. C. Chivers, and J. N. Som, "The acousto-optic interaction in the measurement of ultrasonic transducer surface motion," Ultrasonics **31**(5), 321–325 (1993).

[23] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[24] H. T. O'Neil, "Theory of focusing radiator," J. Acoust. Soc. Am. **21**, 516–526 (1949).

[25] E. L. Madsen, M. M. Goodsitt, and J. A. Zagzebski, "Continuous waves generated by focused radiators," J. Acoust. Soc. Am. **70**(5), 1508–1517 (1981).

[26] A. Pentinnen and M. Luukkala, "The impulse response and pressure nearfield of a curved ultrasonic radiator," J. Phys. D **9**, 1547–1557 (1976).

[27] H. T. O'Nell, "Theory of focusing radiators," J. Acoust. Soc. Am. **21**(5), 516–526 (1949).

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Wang *et al.*: Optical measurement of focused ultrasonic fields   2627

On experimental determination of the random-incidence response of microphones

Salvador Barrera-Figueroa[a)] and Knud Rasmussen[b)]
Danish Primary Laboratory of Acoustics - Danish Fundamental Metrology, Matematiktorvet B307, DK-2800 Lyngby, Denmark

Finn Jacobsen[c)]
Acoustic Technology - Ørsted-DTU, Technical University of Denmark, Ørsteds Plads B352, DK-2800 Lyngby, Denmark

The random-incidence sensitivity of a microphone is defined as the ratio of the output voltage to the sound pressure that would exist at the position of the acoustic center of the microphone in the absence of the microphone in a sound field with incident plane waves coming from all directions. The random-incidence correction of a number of laboratory standard microphones has been determined experimentally. Although the measurement procedure seems to be straightforward, some practical and fundamental problems arise: (i) Reflections from the mounting rig contaminate the measured frequency response, and whereas some of these reflections can be removed using a time-selective technique, others coincide with the direct impulse response and consequently cannot be removed in the time domain and thus affect the accuracy of the estimate; (ii) the accuracy of the estimate is relying on the rotational symmetry of the microphone and depends on the angular resolution. The effect of the angular resolution has been compared with the analytical solution of the scattering and diffraction around a solid sphere. Numerical calculations supplement the experimental results. Although the procedure has only been applied to laboratory standard microphones, it is not restricted to such microphones and may be applied to other types of measurement microphones. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2715655]

I. INTRODUCTION

The sensitivity of a microphone is defined as the ratio of the output voltage to the sound pressure that would exist at the position of the acoustic center of the microphone in the absence of the microphone in a specified sound field. During a calibration, the microphone can be subjected to a uniform sound pressure over the diaphragm, a plane wave in a free field, or a diffuse field. The first two cases have been extensively studied. These activities have led to the development of a number of standards[1–3] while further investigations continue.[4–6] The random-incidence sensitivity has always been considered equivalent to the diffuse-field sensitivity, and used widely in practical applications. The random-incidence sensitivity is defined in terms of the assembled response of the microphone to plane waves from all possible directions impinging successively onto the microphone.[7] However, determining the absolute sensitivity of a microphone at oblique angles of incidence is a very difficult process. Thus, instead of using the free-field sensitivities at each angle of incidence, a practical approach that allows the use of a relative response, such as the directivity factor or the directivity index, can be used with advantage.

The directivity factor can be determined using[8]

$$Q(f) = 4\pi \left[\int_{\Omega} |G(\theta,\phi)|^2 d\Omega \right]^{-1}, \tag{1}$$

where f is the frequency, Ω is the solid angle (4π), and $G(\theta,\phi)$ is the ratio of the frequency response at the angles θ and ϕ to the frequency response at normal incidence, i.e., $G(\theta,\phi) = H(\theta,\phi)/H(0)$. If the microphone can be regarded as rotationally symmetrical, the dependence on ϕ disappears, and the directivity factor can be determined using the following expression:[9]

$$Q(f) = 2 \left[\int_0^\pi |G(\theta)|^2 \sin\theta d\theta \right]^{-1}. \tag{2}$$

This expression can be used when an analytical form of the ratio of responses is available. However, this is not the case when the frequency response of a measurement microphone is determined experimentally. In such a case, the integral in Eq. (2) must be replaced by a discrete series,

$$Q(f) = \frac{2|H(f,\theta_0)|^2}{\sum_{n=1}^{\pi/\Delta\theta} |H(f,\theta_n)|^2 \sin\theta_n \Delta\theta}. \tag{3}$$

Needless to say such a discretization is made under the assumption that it has no significant effect on the final accuracy of the estimated sensitivity. For example, a standard concerned with the determination of the diffuse field calibration of sound level meters specifies that each segment should not be larger than 3% of the total measurement area.[7] The

---

[a)]Electronic mail: sbf@dfm.dtu.dk
[b)]Electronic mail: kr@oersted.dtu.dk
[c)]Electronic mail: fja@oersted.dtu.dk

directivity index is the directivity factor in logarithmic form[8]

$$DI(f) = 10 \log Q(f). \qquad (4)$$

The directivity index can be used for determining the random-incidence sensitivity or the random-incidence correction. The random-incidence sensitivity is determined using

$$M_{RI} = M_{F,\theta=0} - DI(f), \qquad (5)$$

where $M_{F,\theta=0}$ is the free-field sensitivity at normal incidence. The random-incidence correction can be calculated using

$$C_{RI} = C_{F,\theta=0} - DI(f), \qquad (6)$$

where $C_{F,\theta=0}$ is the free-field correction at normal incidence. The free-field correction is the difference between the free-field sensitivity and the pressure sensitivity obtained either by reciprocity calibration or by calibrating using an electrostatic actuator. When a direct measurement of the random-incidence sensitivity is not available, it is common practice to use typical values of the random-incidence correction together with the pressure response of an individual microphone for determining the random-incidence sensitivity of the microphone.

The experimental determination of the directivity index is far from a simple and straightforward process. Measuring the frequency response at all incidence angles is not trivial. Two problems can be identified: (a) The measurements must be carried out in an anechoic environment, and (b) the error introduced by the use of the discrete angular positions around the microphone. Even in the best anechoic rooms the accuracy of the measurement will be degraded by reflections from the walls and from the measurement rig. Hence a technique that can remove such a negative effect is needed. A time-selective technique that has been developed for the reciprocity calibration of microphones in a free field can be applied with advantage.[10]

In this paper an exposition of the experimental determination of the directivity factor and the random-incidence correction of laboratory standard microphones is presented. The application of a time-selective technique for removing unwanted reflections from the frequency response is described. The effect of the angular resolution in the determination of the discrete directivity factor is studied and compared to the case of the scattering by a sphere. Numerical estimates obtained using the boundary element method (BEM) supplement the experimental results.

II. EXPERIMENTAL SETUP

The measurements required for determining the directivity index have to be made in a free field. Usually, this free field is realized in an anechoic room, but a *simulated* free field can also be used. Such a simulated free field is normally the result of time-selective techniques such as time delay spectrometry[11] or FFT-based techniques.[10]

A. Measurement setup

The measurement setup is composed of the measurement rig and the measurement instrumentation. The former
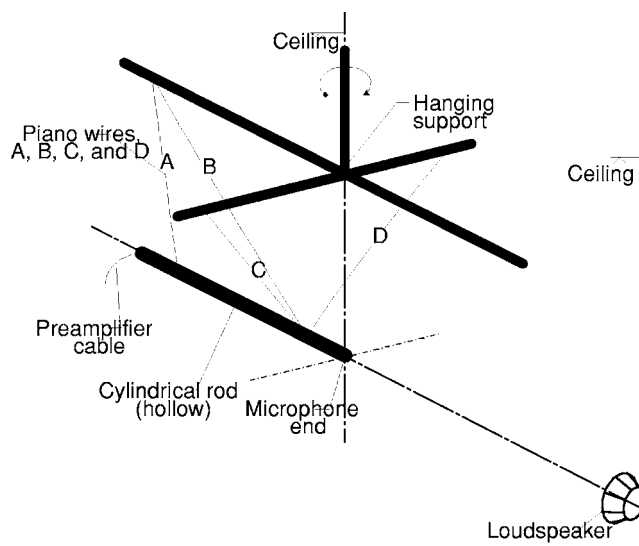
FIG. 1. Schematics of the mounting rig inside the anechoic room. The drawing is not to scale.

was mounted inside the anechoic room and the latter was placed in a control room. The measurement rig has to comply with two criteria: (i) It has to be as reflection-free as possible, and (ii) the rod where the microphones are mounted has to be long enough to be considered semi-infinite and it should have the same cross section as the microphone. A measurement rig was built complying with these two criteria. A scheme and a photograph of the measurement rig are shown in Figs. 1 and 2.

FIG. 2. Photograph of the experimental rig in the anechoic chamber.
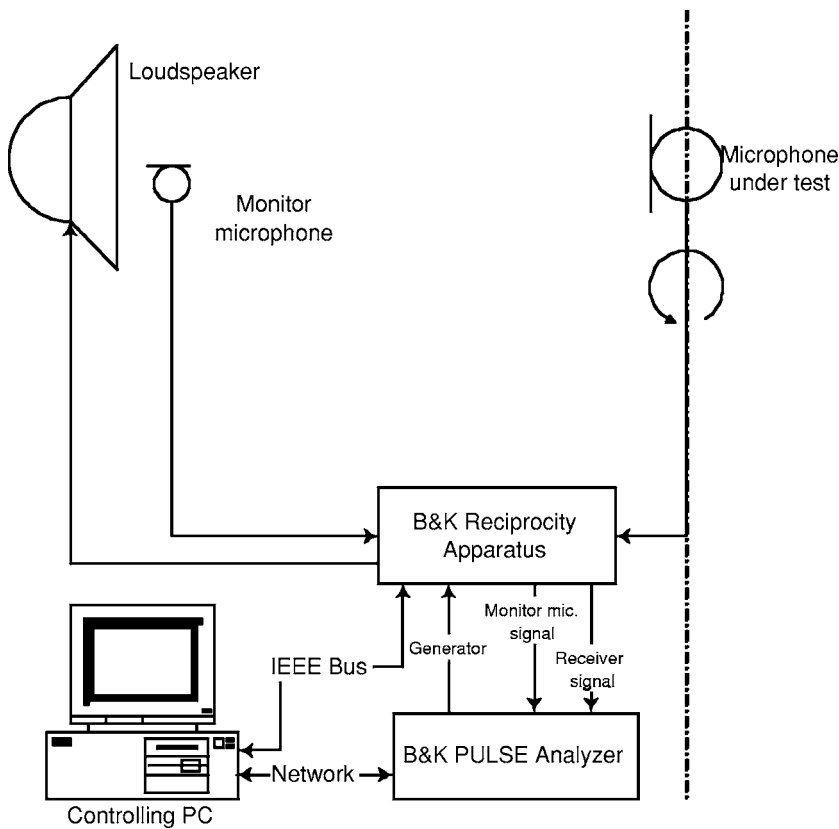
FIG. 3. Schematics of the measurement system used in the measurement of the directivity factor.

The rig is in the middle of a large anechoic room with free volume of nearly 1000 m^3. The rod where the microphones are mounted is about 80 cm long, and it has a diameter of 23.77 or 12.7 mm depending on whether the microphones are 1-in. laboratory standard microphones (LS1) or $\frac{1}{2}$-in. laboratory standard microphones (LS2). The cross that hangs from the ceiling is about 1 m away from the microphone rod, and the piano wires are less than 1 mm in diameter. The microphone and microphone rod are aligned in such a way that the rig rotates around the diaphragm of the microphone. The loudspeaker is 1.7 m away from the microphone. A monitor microphone is placed in front of the loudspeaker.

The microphone under test and the monitor microphone are connected to a Brüel & Kjær (B&K) reciprocity apparatus model 5998. The output of the reciprocity apparatus is connected to a B&K "PULSE" analyzer. The reciprocity apparatus is used for measuring the insert voltage in the channel of the microphone under test. The generator output of the PULSE analyzer is connected to a power amplifier that feeds the loudspeaker. The loudspeaker is a modified tweeter that has a flat frequency response up to 40 kHz. The monitor microphone is a $\frac{1}{4}$-in. condenser microphone, B&K 4136. The analyzer and the reciprocity apparatus are connected to a PC that controls the measurements. Figure 3 shows the measurement system. Determining the frequency response of the test microphone at a given angle, $H(\theta)$, involves measuring the output voltage of the monitor and the test microphone. The measured frequency response is defined as the ratio of the open-circuit voltage of the microphone under test to the output voltage of the monitor microphone, $H_{meas}(f,\theta) = U_{test}(f,\theta)/U_{monitor}(f)$. The frequency response is measured at discrete frequencies using stepped-sine excitation with 30 Hz step in the interval from 1 to 20 kHz and from 2 to 30 kHz for LS1 and LS2 microphones, respectively. Measurements below 1 kHz were not made because of a high level of distortion of the loudspeaker.

B. Removal of reflections from the rig

The measurement rig has been carefully designed so as to minimize reflections that could contaminate the measured frequency response. However, many elements of the rig will reflect sound back to the microphones. In most cases these unwanted reflections could be removed using a time-selective technique. This technique has been tried before in free-field reciprocity calibrations.[10]

The procedure applied in this case was very similar: (a) The frequency response was measured as the transfer function between the signals of two microphones in a frequency interval (f_0, f_{max}); (b) the missing lower portion of the frequency response $(0, f_0)$ was calculated using theoretical data; (c) the high frequency response was taken smoothly to zero by applying a low-pass filter; (d) an inverse Fourier transform was calculated; (e) a time-selective window was applied to the resulting impulse response in order to remove reflections; and (f) a Fourier transform was applied to the "cleaned" impulse response.

If an inverse Fourier transform is to be applied onto the frequency response, this has to be defined in the whole frequency interval, i.e., from 0 to ∞ for a one-sided frequency response. The frequency response is completed following steps (b) and (c). Thus, the frequency response after step (b) can be defined as

Barrera-Figueroa *et al.*: Random-incidence response of microphones
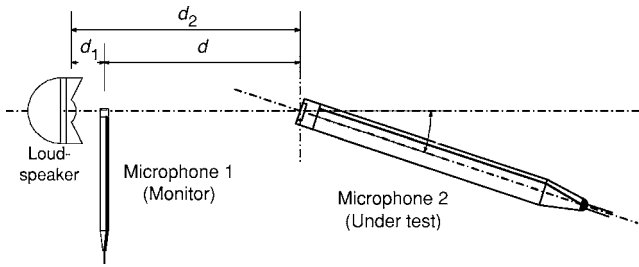
FIG. 4. Schematics of the relative positions of the microphones during the measurements (the drawing is not to scale).

$$H_c(f,\theta) = \begin{cases} H'(f,\theta), & 0 \le f \le f_0 \\ H_{\text{meas}}(f,\theta), & f_0 \le f \le f_{\text{max}}, \end{cases} \quad (7)$$

where $H'(f,\theta)$ can be estimated using the sensitivities and the distance between monitor and test microphones as follows:

$$H'(f,\theta) = \frac{p_2 M_{F,2}(f,\theta)}{p_1 M_{F,1}(f,\theta)}, \quad (8)$$

where p_1 and p_2 are the sound pressures at the position of microphones 1 and 2; and $M_{F,1}(f,\theta)$, and $M_{F,2}(f,\theta)$ and the free-field sensitivities of microphones 1 and 2 at the angle θ and the frequency f. Assuming that the source is a point source, and substituting the sound pressures at points 1 and 2, this expression becomes

$$H'(f,\theta) = \frac{M_{F,2}(f,\theta)}{M_{F,1}(f,\theta)} \frac{d_2}{d_2 + d} \exp(-ikd), \quad (9)$$

where $d = d_2 - d_1$ is the distance between microphones, d_1 and d_2 are the distances from the source to microphones 1 and 2, and k is the wave number. In this case, microphone 1 was the monitor microphone, which remained at a fixed position in front of the loudspeaker while microphone 2, the microphone under test, was exposed to sound from a different angle of incidence at each measurement; a scheme is shown in Fig. 4. Thus, the resulting frequency response is basically the ratio of the sensitivities of the two microphones and the effect of the distance between the two microphones. Pressure sensitivities were used instead of the actual free-field sensitivities in Eq. (9), because at low frequencies there is little difference between the free-field sensitivity and the pressure sensitivity. In practice, the monitor microphone was located close to the source, and therefore interference caused by multiple reflections from the loudspeaker with the direct sound field may be expected. However, such effects are judged to be negligible in the frequency range of concern. In principle, the distance should take account of the positions of the acoustic centers of the microphones and the loudspeaker.[7] However, as the distance between the microphone under test and the loudspeaker was about 1.7 m, the effect of the acoustic center may be considered negligible.

The frequency response of a microphone will decay at high frequencies because of the fact that above the resonance frequency of the diaphragm, its movement is controlled by its mass, and thus the sensitivity of the microphone approaches zero. Therefore such a decay can be artificially made in the measured frequency range, provided that the upper frequency is at least two times the resonance frequency of the microphones. The decay can be accelerated by applying a low-pass filter. In this case, the low-pass filter was a linear phase finite impulse response (FIR) filter of 85th order, with a normalized passband frequency of 0.9, and a stop-band attenuation of 80 dB. The pass-band ripple was 0.0001 dB. The frequency response after operation (c) is

$$H_L(f,\theta) = H_c(f,\theta)L(f), \quad (10)$$

where $L(f)$ is the low-pass filter.

The impulse response was obtained after applying an inverse Fourier transform to $H_L(f,\theta)$ [step (d)]. The time-selective window applied to the impulse response [step (e)] was a Tukey window with a duration of 3 ms. The smoothing percentage of the sides of the window was 70%. The window was long enough to eliminate all the reflections from the measurement rig, and the tapering reduces the ripple at the extremes of the frequency response. The length of the time selective window was defined from two criteria: (a) The distance between microphones, and (b) a visual inspection of the impulse response. The first criterion is based on the fact that the distance between microphones defines the instants when reflections from walls and other interferences occur. The second criterion is a visual inspection that ensures that no portion of the direct impulse response is excluded, and unwanted effects in as far as possible are excluded.

In some cases, the reflections could not be separated from the impulse response. This occurred at incidence angles larger than 120°. This is because perturbations from any element behind the microphone will be a part of the direct wave. Therefore, it is very important to minimize reflections coming from any element behind the microphone. For instance, the setup originally had a vertical rod that supported the horizontal rod where the microphone is located. The reflections from the vertical rod were significant, and they could not be removed from the measurement. Because of this, it was decided to modify the rig. The major change was the substitution of the vertical rod by piano wires. As can be seen in Fig. 5, only the reflection from the hanging preamplifier cable remained. This demonstrates a fundamental problem of the method: Unless the cable is removed, there is no easy way to get rid of such a reflection or of reflections of any element behind the microphone such as the piano wires and the mounting elements.

In spite of this, the time-selective technique proved to be useful in eliminating other strong reflections from elements of the setup, such as a grid-floor and the flat platform below the microphones that was used to place the angular divider. An illustration of how large the influence of such reflections for LSI microphones was is shown in Fig. 6. Figure 6(a) shows the difference between the "cleaned" and "raw" frequency responses for a measurement at $\theta = 0°$. It can be seen that at such an angle the reflections from fixed elements of the measurement rig (platform, hanging structures, etc.) are significant, up to 1 dB below 5 kHz, while the elements behind the microphone have a small influence. Figure 6(b) shows the same results for a measurement at $\theta = 135°$. It is clear in this figure that the elements behind the microphone now have a significant influence.
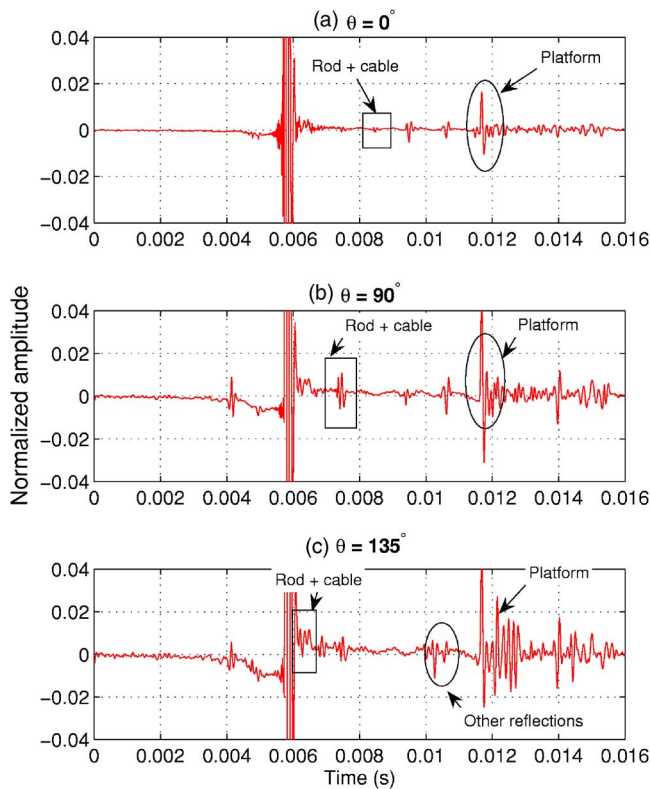
FIG. 5. Impulse response of the modified rig at different angles of incidence.

FIG. 6. Illustration of the effect of the cleaning technique: (a) Measurement of the frequency response at $\theta=0°$, and (b) measurement of the frequency response at $\theta=135°$. In the two graphs, the solid line is the frequency response with reflections removed, the dotted line is the measured frequency response, and the dashed line is the difference between the measured and the "cleaned" frequency responses.

III. THE INFLUENCE OF THE ANGULAR RESOLUTION

The use of Eq. (2) would yield the exact value of the directivity factor. However, such an expression cannot be used in the case of the microphones. Instead, the directivity factor must be obtained using the approximation given by Eq. (3). Therefore, the influence of the angular step on the final estimate of the quantity sensitivity should be analyzed.

A suitable procedure would consist in determining the directivity factor using different angular resolutions and examining whether the estimate converges to a unique value as the resolution gets finer. Although it is intuitively evident that the finer the resolution the better the result, there is no actual indication of how accurate the estimate determined from the finest resolution will be. On the other hand, Eq. (1) can be implemented for a case where the analytical solution of the scattering problem exists, such as the case of a sphere. Once this estimate is obtained it can be compared with the outcome of the implementation of Eq. (3) for the same case. The analytical and the discrete estimates are analyzed in the following.

A. The case of a sphere

If harmonic variation with the $\exp(-i\omega t)$ sign convention is assumed, the sound pressure scattered by a solid sphere of radius a, centered at the origin of a spherical coordinate system, can be calculated from[12]

$$p_s(r,\theta) = -A\sum_m \{(2m+1)i^{m+1}$$

$$\times \exp(-i\delta_m)\sin(\delta_m)P_m(\cos\theta)h_m(kr)\}, \quad (11)$$

where A is the amplitude of an incident plane wave coming from the direction θ, $h_m(x)$ is the spherical Hankel function of the first kind and order m, P_m is the Legendre function of order m, k is the wave number, r is the distance to the observation point which in this case is the radius of the sphere, and the angle δ_m is defined as

$$\delta = \arctan\left\{\frac{(m+1)j_{m+1}(ka) - mj_{m-1}(ka)}{mn_{m-1}(ka) - (m+1)n_{m+1}(ka)}\right\}, \quad (12)$$

where j_m and n_m are spherical Bessel and Neumann functions of order m.

In order to determine the total sound pressure the sound pressure of the incident plane wave must be added. In spherical coordinates this is

$$p_i(r,\theta) = A\sum_{m=0}^{\infty} (2m+1)i^{m+1}P_m(\cos\theta)j_m(kr). \quad (13)$$

Thus, the total pressure on the surface of the sphere is

$$p_t(r,\theta) = A\left\{\sum_{m=0}^{\infty} (2m+1)i^m P_m(\cos\theta)[j_m(ka) - i\right.$$

$$\left.\times \exp(-i\delta_m)\sin(\delta_m)j_m(ka)]\right\}. \quad (14)$$

Equation (14) can be then integrated as in Eq. (1) or it can be
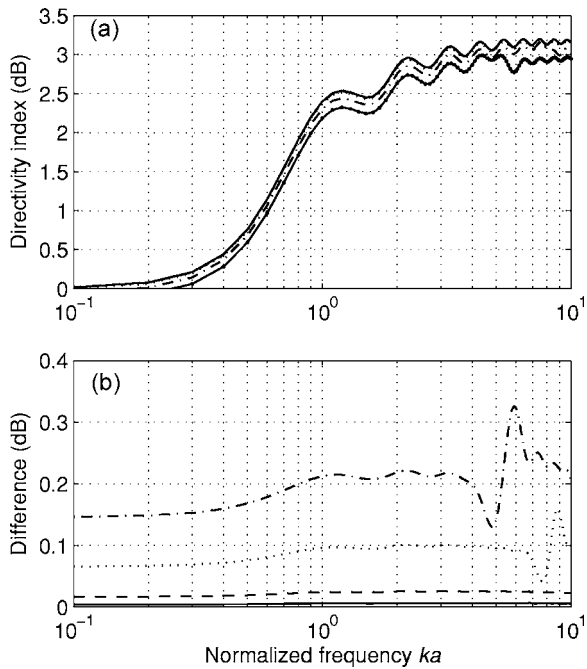
FIG. 7. (a) Directivity index obtained: analytically (solid line), and from discrete values of θ, 5° (dashed line), 10° (dotted line), 20° (dashed-dotted line), and 30° (solid line with dot markers). (b) Difference between the analytical and the discrete estimate of the directivity index: 5° (solid line), 10° (dashed line), 20° (dotted line), and 30° (dashed-dotted line).

FIG. 8. (a) Directivity index of LS1 microphones obtained experimentally using discrete values of θ, 5° (solid line), 10° (dashed line), 20° (dotted line), and 30° (dashed-dotted line), (b) Difference between the discrete estimate of the directivity index using 5° and the estimate obtained using 10° (solid line), 20° (dashed line), and 30° (dotted line). The solid lines with open-circle, squared and closed-circle markers are the numerical estimates of the difference.

estimated at discrete values of θ, and then an estimate of the directivity factor can be determined using Eq. (3).

Figure 7 shows the difference between an estimate of the directivity index obtained by numerical integration and estimates obtained from the discrete summation of the function at discrete values of θ, 20°, 10°, and 5°. The results are discussed further in Sec. V.

B. The case of a microphone

Unlike the case of the sphere, there is no analytical expression for the diffraction of an impinging sound wave on a microphone. Therefore, it is only possible to apply Eq. (3) to the ratio of the frequency response measured at discrete values of the angle θ to the frequency response at $\theta = 0°$. Thus, it is not possible to establish an exact reference of the directivity factor that can be compared with the discrete estimate. However, the differences between discrete estimates obtained using increasing angular steps can be compared with the case of the sphere. Figure 8 shows the difference between discrete estimates obtained using 5°, 10°, 20°, and 30°. The results are discussed further in Sec. V.

IV. DIRECTIVITY INDEX AND RANDOM-INCIDENCE CORRECTION

The directivity index of LS1 and LS2 microphones has been determined experimentally. Four LS1 microphones were used in the investigation, and one of them was measured four times whereas another was measured twice. This gives a total of eight measurements. The results are shown in Fig. 9(a). Six LS2 microphones were used in the investiga-

tion as well. Two of them were measured twice; this gives a total of eight measurements. The results are shown in Fig. 9(b).

Equations (5) and (6) give the possibility of estimating the random-incidence sensitivity or the random incidence correction from the directivity factor and the free-field sensitivity or the free-field correction of the microphone at $\theta = 0$. Figure 10 shows the random-incidence correction of LS1 and LS2 microphones obtained experimentally and numerically. Tables I and II list the values of the random-incidence correction and the standard deviation of the measurements of the directivity index. The estimate of the random-incidence correction has been obtained using Eq. (6). The free-field correction is the difference between free-field and pressure sensitivities as described in Ref. 3.

The numerical estimates were obtained using the BEM formulation described in Ref. 13. The geometry used in the simulations was a microphone mounted on an end of a cylindrical rod of the same diameter of the microphones. The semi-infinite rod was approximated by a rod of 60 cm long with a hemispherical back end. This will introduce a small disturbance in the simulated results because of reflections from the back of the rod. However, because of the length of the rod, these reflections are expected to have a small amplitude. The frequency range used in the calculations is from 1 to 20 kHz for LS1 microphones and from 1 to 30 kHz for LS2 microphones. The size of the smallest element in the axisymmetric mesh is 2.5 and 1.5 mm for LS1 and LS2 microphones, respectively. Thus, there will be at least 4 ele-
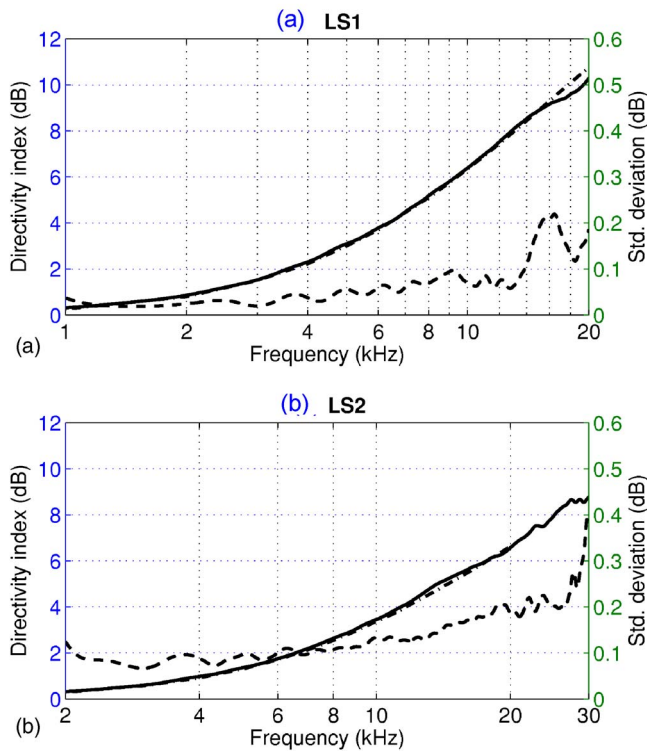
FIG. 9. Experimental directivity index of (a) LS1, and (b) LS2 microphones: Average (solid line), standard deviation (dashed line), and numerical estimate (dashed-dotted line). Standard deviation scale on right y axis.

FIG. 10. Random-incidence correction of (a) LS1, and (b) LS2 microphones: Experimental (solid line), and numerical estimate (solid line with dot markers).

ments per wavelength at the highest frequency. In order to avoid the nonuniqueness problem a random CHIEF point has been added in the interior of the geometry as described in Ref. 14, and the calculation have been checked by determining the condition numbers of the BEM matrices[15] and by repeating calculations with small frequency shifts. In this investigation it has been assumed that the displacement of the microphone diaphragm is described by a Bessel function of the first kind.[12,16]

V. DISCUSSION

The time-selective technique is very effective in removing reflections from the measurement rig, especially those that are fully separated from the direct impulse response. The rig can be optimized for reducing the amplitude of the reflections. However, even when a time-selective technique is used some reflections will coincide with a part of the direct impulse response.

The difference between the analytical and the discrete estimates of the directivity index of the sphere is nearly constant in the whole frequency range for all angular steps, whereas the directivity index of the microphone increases with the frequency, and the slope changes as the angular resolution coarsens. The reason is that microphones are much more directional than spheres. However, the differences between the discrete estimates obtained using 5° and 10° are of the same order of magnitude for the microphone and the sphere, about 0.01 dB at low frequencies; this is also the case for the difference between 10° and 30°, about

0.06 dB at low frequencies. Thus, the difference between the analytical estimate and the discrete estimate for the sphere gives a reliable idea of the accuracy that can be reached using different angular resolutions. A resolution of 5° seems to be adequate because the difference is less than 0.01 dB; using a resolution finer than 5° would improve the accuracy very little while increasing the measurement time because of the additional measurements needed. The numerical estimate of the differences confirms the experimental results.

TABLE I. Random-incidence correction and the standard deviation of the directivity index of LS1 microphones.

f (kHz)	C_{RI} (dB)	s (dB)	f (kHz)	C_{RI} (dB)	s (dB)
1.0	−0.05	0.04	8.5	3.41	0.08
1.5	0.00	0.02	9.0	3.37	0.09
2.0	0.13	0.02	9.5	3.23	0.09
2.5	0.31	0.03	10.0	3.06	0.07
3.0	0.60	0.02	10.5	2.83	0.06
3.5	0.86	0.04	11.0	2.56	0.09
4.0	1.25	0.04	11.5	2.27	0.06
4.5	1.60	0.04	12.0	1.88	0.07
5.0	2.00	0.06	12.5	1.45	0.06
5.5	2.43	0.04	13.0	0.94	0.06
6.0	2.76	0.06	13.5	0.40	0.08
6.5	3.10	0.05	14.0	−0.17	0.11
7.0	3.27	0.06	14.5	−0.70	0.15
7.5	3.40	0.07	15.0	−1.03	0.19
8.0	3.43	0.07			

Barrera-Figueroa *et al.*: Random-incidence response of microphones

TABLE II. Random-incidence correction and the standard deviation of the directivity index of LS2 microphones.

f (kHz)	C_{RI} (dB)	s (dB)	f (kHz)	C_{RI} (dB)	s (dB)
2.0	−0.05	0.12	16.5	3.06	0.17
2.5	−0.04	0.08	17.0	3.08	0.18
3.0	0.01	0.07	17.5	3.05	0.18
3.5	0.03	0.09	18.0	2.96	0.18
4.0	0.08	0.09	18.5	2.91	0.20
4.5	0.17	0.08	19.0	2.87	0.20
5.0	0.26	0.10	19.5	2.75	0.20
5.5	0.39	0.08	20.0	2.57	0.19
6.0	0.51	0.10	20.5	2.37	0.19
6.5	0.64	0.11	21.0	2.19	0.18
7.0	0.78	0.10	21.5	2.05	0.20
7.5	0.94	0.11	22.0	1.84	0.22
8.0	1.10	0.11	22.5	1.58	0.20
8.5	1.30	0.11	23.0	1.42	0.20
9.0	1.48	0.11	23.5	1.41	0.22
9.5	1.64	0.12	24.0	1.28	0.22
10.0	1.85	0.13	24.5	1.03	0.20
10.5	2.05	0.13	25.0	0.80	0.20
11.0	2.20	0.13	25.5	0.55	0.20
11.5	2.37	0.13	26.0	0.34	0.20
12.0	2.51	0.13	26.5	0.16	0.20
12.5	2.60	0.13	27.0	−0.04	0.22
13.0	2.68	0.13	27.5	−0.24	0.26
13.5	2.74	0.14	28.0	−0.21	0.26
14.0	2.83	0.15	28.5	−0.39	0.26
14.5	2.93	0.16	29.0	−0.43	0.30
15.0	2.99	0.16	29.5	−0.52	0.35
15.5	3.04	0.17	30.0	−0.80	0.40
16.0	3.07	0.17			

An alternative might be to distribute the angles of incidence uniformly over a solid angle of 4π such that their weighting is all the same. However, this will not improve the deviations observed at frequencies where the microphones are more directional because the number of angular positions close to the normal incidence is sparse compared to the number of positions around grazing incidence (90°). This would result in undersampling the area where the directivity of the microphones is most significant.

The reproducibility of the directivity index is better for LS1 microphones than for LS2 microphones. This is not unexpected because LS2 microphones have a lower sensitivity. The good behavior of LS1 microphones makes it possible to observe a significant deviation between 14 and 18 kHz, reaching a maximum of about 16 kHz. It is known that around 16 kHz there is a resonance in the back cavity of the LS1 microphones. This resonance may be excited in different ways depending on how the oblique incidence modifies the movement of the diaphragm. Such a behavior cannot be detected in the LS2 case. The reason may be that the diaphragm of the LS2 is less compliant and more damped than the diaphragm of the LS1 microphones.

The estimation of the directivity index is a necessary step for determining the random-incidence response unless the absolute response of the microphones at any angle of incidence is available. The accuracy of the estimate of the random-incidence correction will be a combination of the accuracy of the experimental estimate of the directivity index and the accuracy of the normal-incidence response. Thus, the most accurate estimate of the random-incidence of a microphone can be obtained using the individual normal-incidence correction for a particular microphone. The agreement between the experimental and the numerical estimates of the random-incidence correction is very good at frequencies below the resonance frequency of each type of microphone. Above resonance the agreement degrades because of the assumed velocity distribution of the microphone membrane. The upper frequency is limited by the validity of the polynomial approximation used for calculating the normal-incidence free-field correction (12.5 kHz for LS1 and 25 kHz for LS2).

Finally, it is worth mentioning that the procedure for measuring the response of an axially symmetrical microphone can easily be extended to cover the case of microphones with no axial symmetry. In such a case, a discretized version of Eq. (1) should be used instead.

VI. CONCLUSIONS

The directivity index and the random-incidence correction of laboratory standard microphones have been determined experimentally. Unwanted reflections from the measurement rig were removed from the frequency responses using a FFT-based, time-selective procedure.

The influence of the angular resolution has been studied by determining the directivity index using different angular resolutions. The case of the diffraction of the microphone was compared with the case of the diffraction of a sphere. The comparison showed that a resolution of 5° represents a good compromise between accuracy and measurement time.

The accuracy of the random-incidence response depends on the accuracy of the estimate of the directivity index and the estimate of the normal-incidence free-field response. Although standardized values can be used, the lowest uncertainty is obtained when individual values of the free-field correction having a lower uncertainty than the standardized values are used for a particular microphone.

[1] IEC International Standard 61094-2, "Measurement microphones Part 2: Primary method for pressure calibration of laboratory standard microphones by the reciprocity technique" (International Electrotechnical Commission, Geneva, 1995).

[2] IEC International Standard 61094-3, "Measurement microphone Part 3: Primary method for free-field calibration of laboratory standard microphones by the reciprocity technique" (International Electrotechnical Commission, Geneva, 1995).

[3] IEC Technical Specification 61094-7, "Measurement microphones Part 7: Values for the difference between free-field and pressure sensitivity levels of laboratory standard microphones" (International Electrotechnical Commission, Geneva, 2006).

[4] C. Guianvarc'h, J.-N. Durocher, M. Bruneau, and A.-M. Bruneau, "Improved formulation of the acoustic transfer admittance of cylindrical cavities," Acust. Acta Acust. 92, 345–354 (2006).

[5] R. A. Kosobrodov and S. V. Kuznetsov, "Acoustic transfer impedance of plane-wave coupler," Acust. Acta Acust. 92, 513–520 (2006).

[6] S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, "On experimental determination of the free-field correction of Laboratory Standard microphones at normal incidence," Metrologia 44, 57–63 (2007).

[7] IEC International Standard 61183, "Random incidence and diffuse-field calibration of sound level meters" (International Electrotechnical Commis-

sion, Geneva, 1994).

[8] L. E. Kinsler, A. R. Frey, A. B. Copens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 2000).

[9] L. L. Beranek, *Acoustical Measurements* (Acoustical Society of America, Melville, 1988).

[10] S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, "A time-selective technique for free-field reciprocity calibration of condenser microphones," J. Acoust. Soc. Am. **114**, 1467–1476 (2003).

[11] R. Heyser, "Acoustical measurements by time delay spectrometry," J. Audio Eng. Soc. **15**, 370–382 (1967).

[12] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968/1984).

[13] P. M. Juhl, "A numerical investigation of standard condenser microphones," J. Sound Vib. **177**, 433–446 (1994).

[14] H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 45–58 (1968).

[15] P. M. Juhl, F. Jacobsen, V. Cutanda-Henríquez, and S. Quirós-Alpera, "On the non-uniqueness problem in a 2-D half-space BEM formulation," in *Proceedings of Ninth International Congress on Sound and Vibration*, Orlando, FL (2002), CD-ROM.

[16] A. J. Zuckerwar, "Theoretical response of condenser microphones," J. Acoust. Soc. Am. **64**, 1278–1285 (1978).

Extracting the Green's function of attenuating heterogeneous acoustic media from uncorrelated waves

Roel Snieder[a)]

Center for Wave Phenomena and Department of Geophysics, Colorado School of Mines, Golden Colorado 80401

The Green's function of acoustic or elastic wave propagation can, for loss-less media, be retrieved by correlating the wave field that is excited by random sources and is recorded at two locations. Here the generalization of this idea to attenuating acoustic waves in an inhomogeneous medium is addressed, and it is shown that the Green's function can be retrieved from waves that are excited throughout the volume by spatially uncorrelated injection sources with a power spectrum that is proportional to the local dissipation rate. For a finite volume, one needs both volume sources and sources at the bounding surface for the extraction of the Green's functions. For the special case of a homogeneous attenuating medium defined over a finite volume, the phase and geometrical spreading of the Green's function is correctly retrieved when the volume sources are ignored, but the attenuation is not. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2713673]

I. INTRODUCTION

The extraction of the Green's function by correlating waves excited by random sources that are recorded at two locations has recently received much attention. There are numerous derivations of this principle that are valid for closed systems[1] and for open systems (e.g., Refs. 2–4). Formulations of this principle are based either on random sources placed throughout a volume[1,5] or on sources that are located at a surface.[6–8] The extraction of the Green's function using random wave fields has been applied to ultrasound,[9–12] in seismic exploration,[13–15] in crustal seismology,[16–19] in ocean acoustics,[20–22] to buildings,[23,24] and in helioseismology.[25–27] The recent supplement of seismic interferometry[28] in Geophysics gives an overview of this field of research. Phrases that include passive imaging, correlation of ambient noise, extraction of the Green's function, and seismic interferometry have been proposed for this line of research. Recently the theory has been developed for the extraction of the Green's function for more general linear systems than acoustic or elastic waves.[29,30]

Many derivations of this principle are valid for systems that are invariant under time reversal. Several derivations invoke time-reversal invariance explicitly.[2,3,13] For acoustics waves in a flowing medium the time-reversal invariance is broken by the flow; this broken symmetry has been incorporated in the theory for the extraction of the Green's function.[31,32] Attenuation also breaks the invariance for time reversal. For homogeneous acoustic media[5,33] and for a homogeneous oceanic waveguide[21] attenuation has been incorporated into the theory for the extraction of the Green's function. Weaver and Lobkis[4] use complex frequency as a tool to force convergence on an integral over all sources.

Here I derive the principle of seismic interferometry for general attenuating, acoustic media, and extend earlier for-

mulations for homogeneous media to include arbitrary heterogeneity in density, compressibility, and intrinsic attenuation. Section II introduces the basic equations and rederives a representation theorem of the correlation type for attenuating media. Section III shows that for an unbounded volume, or for a volume that is bounded by a surface where the pressure or normal component of the velocity vanishes, the Green's function can be extracted from waves excited by uncorrelated volume sources with a source strength that is proportional to the local dissipation rate. Section IV shows that for a bounded volume one needs, in general, both volume sources and surface sources in order to retrieve the correct Green's function. Section V illustrates the relative roles of the volume sources and surface sources by analyzing the special case of a homogeneous, attenuating medium, with a single reflector. In this special case, when volume sources are ignored, the phase and geometrical spreading of the Green's function are correctly reproduced by seismic interferometry, but the attenuation is not.

II. BASIC EQUATION FOR ACOUSTIC WAVES

Using the Fourier convention $f(t) = \int f(\omega) \exp(-i\omega t) d\omega$, the pressure p and particle velocity $\mathbf{v}$ for acoustic waves satisfy, in the frequency domain, the following coupled equations:

$$\nabla p - i\omega\rho\mathbf{v} = 0, \tag{1}$$

$$(\nabla \cdot \mathbf{v}) - i\omega\kappa p = q. \tag{2}$$

In these expressions ω is the angular frequency, ρ the mass density, and κ the compressibility. All expressions in this work are given in the frequency domain; for brevity this frequency-dependence is not denoted explicitly. It is assumed that only injection sources q are present. Body forces would render the right-hand side of expression (1) nonzero. For attenuating media, the compressibility κ is complex, this

---

[a)]Electronic mail: rsnieder@mines.edu

quantity can be decomposed in a real and imaginary part:

$$\kappa = \kappa_r(\mathbf{r}, \omega) + i\kappa_i(\mathbf{r}, \omega). \tag{3}$$

Because of the Kramers-Kronig relation (e.g., Refs. 34 and 35), the real and imaginary parts of the compressibility depend on frequency. In contrast to the treatment of de Hoop,[36] it is presumed that the mass density is real. In this general derivation the density and compressibility can be arbitrary functions of location and frequency.

Following de Hoop[37] and Fokkema and van den Berg,[38] expressions (1) and (2) can be used to derive a representation theorem of the correlation type. The treatment given here generalizes earlier descriptions of the extraction of the Green's function[7,32] to include dissipation. Two wave states, labeled A and B, are considered that both satisfy expressions (1) and (2), and that are excited by forcing functions q_A and q_B, respectively. The subscripts A and B indicate the state for each quantity. A representation theorem of the correlation type is obtained by integrating the combination $(1)_A \cdot \mathbf{v}_B^* + (1)_B^* \cdot \mathbf{v}_A + (2)_A p_B^* + (2)_B^* p_A$ over volume, and applying Gauss' theorem. [The asterisk denotes complex conjugation, and $(1)_B^*$ stands, for example, for the complex conjugate of expression (1) for state B.] This gives

$$\oint (p_A \mathbf{v}_B^* + p_B^* \mathbf{v}_A) \cdot d\mathbf{S} = \int (q_B^* p_A + q_A p_B^*) dV - i\omega$$
$$\times \int (\kappa^* - \kappa) p_A p_B^* dV, \tag{4}$$

where $\oint (\cdots) \cdot d\mathbf{S}$ denotes the surface integral over the surface that bounds the volume. Note that the last term is due to the attenuation; for loss-less media κ is real, and $\kappa^* - \kappa = 0$. The relative roles of the surface integral on the left-hand side and the volume integral in the last term play a crucial role in the following treatment. In the following the "surface" refers to the surface that bounds the volume. In the presence of cavities this surface may consist of disconnected pieces.

These representation theorems can be used to derive several properties of the Green's function $G(\mathbf{r}, \mathbf{r}_0)$ that is the pressure response to an injection source $q(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_0)$. Setting

$$q_{A,B}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_{A,B}) \tag{5}$$

implies that the corresponding pressure states are given by

$$p_{A,B}(\mathbf{r}) = G(\mathbf{r}, \mathbf{r}_{A,B}), \tag{6}$$

respectively.

Inserting the excitations (5) into expression (4), and using Eq. (1) to eliminate the velocity, one obtains

$$G^*(\mathbf{r}_A, \mathbf{r}_B) + G(\mathbf{r}_B, \mathbf{r}_A)$$
$$= 2\omega \int \kappa_i(\mathbf{r}, \omega) G(\mathbf{r}, \mathbf{r}_A) G^*(\mathbf{r}, \mathbf{r}_B) dV$$
$$+ \oint \frac{1}{i\omega\rho} (G^*(\mathbf{r}, \mathbf{r}_B) \nabla G(\mathbf{r}, \mathbf{r}_A)$$
$$- G(\mathbf{r}, \mathbf{r}_A) \nabla G^*(\mathbf{r}, \mathbf{r}_B)) \cdot d\mathbf{S}. \tag{7}$$

For brevity the frequency-dependence of G is suppressed. In

the presence of intrinsic attenuation, reciprocity of acoustic waves still holds, hence

$$G(\mathbf{r}_A, \mathbf{r}_B) = G(\mathbf{r}_B, \mathbf{r}_A). \tag{8}$$

Expression (7) can therefore be written as

$$G^*(\mathbf{r}_B, \mathbf{r}_A) + G(\mathbf{r}_B, \mathbf{r}_A)$$
$$= 2\omega \int \kappa_i(\mathbf{r}, \omega) G(\mathbf{r}_A, \mathbf{r}) G^*(\mathbf{r}_B, \mathbf{r}) dV$$
$$+ \oint \frac{1}{i\omega\rho} (G^*(\mathbf{r}_B, \mathbf{r}) \nabla G(\mathbf{r}_A, \mathbf{r})$$
$$- G(\mathbf{r}_A, \mathbf{r}) \nabla G^*(\mathbf{r}_B, \mathbf{r})) \cdot d\mathbf{S}. \tag{9}$$

Note that for loss-less media, because of the complex conjugates, the surface integral does not vanish when the system satisfies radiation boundary conditions at the surface. A similar relation has been derived for electromagnetic fields in conducting media.[39]

The left-hand side of expression (9) is the sum of the causal and acausal Green's functions. Wapenaar et al.[7] use this expression for loss-less media ($\kappa_i = 0$) to show that the sum of the causal and acausal Green's function can be obtained by cross-correlating the pressure fields that are due to uncorrelated random sources at the surface. The pressure field caused by these sources is transmitted to the points $\mathbf{r}_A$ and $\mathbf{r}_B$ in the interior by the Green's functions in the surface integral in Eq. (9). For attenuating media $[\kappa_i(\mathbf{r}) \neq 0]$, this analysis is complicated by the presence of the volume integral in this expression.

III. INTERFEROMETRY WHEN THE SURFACE INTEGRAL VANISHES

This section analyzes the special case where the surface integral in expression (9) vanishes. This is the case when one of the following conditions is satisfied:

C1: The volume integration is over all space. For attenuating media the wave field vanishes exponentially at infinity, and the surface integral vanishes.
C2: The pressure vanishes at the surface ($G = 0$).
C3: The normal component of the velocity perpendicular to the surface vanishes at the surface. Because of expression (1) this implies that $\nabla G \cdot d\mathbf{S} = 0$.

When one of the conditions C1–C3 is satisfied, the pressure is related to the excitation by

$$p(\mathbf{r}_0) = \int G(\mathbf{r}_0, \mathbf{r}) q(\mathbf{r}) dV \tag{10}$$

and the representation theorem of the correlation type (9) reduces to

$$G^*(\mathbf{r}_B, \mathbf{r}_A) + G(\mathbf{r}_B, \mathbf{r}_A) = 2\omega \int \kappa_i(\mathbf{r}, \omega) G(\mathbf{r}_A, \mathbf{r}) G^*(\mathbf{r}_B, \mathbf{r}) dV. \tag{11}$$

Consider the situation where random pressure sources are present throughout the volume, and that these sources at different locations are uncorrelated:

$$q(\mathbf{r}_1, \omega) q^*(\mathbf{r}_2, \omega) = \kappa_i(\mathbf{r}_1, \omega) \delta(\mathbf{r}_1 - \mathbf{r}_2) |S(\omega)|^2, \tag{12}$$

where $|S(\omega)|^2$ denotes the power spectrum of the random excitation. The excitation (12) is proportional to $\kappa_i(\mathbf{r}, \omega)$, the imaginary part of the compressibility, which in turn is proportional to the local attenuation. This means that the excitation (12) supplies a random excitation of the pressure field that locally compensates for the attenuation. Multiplying expression (11) with $|S(\omega)|^2$ gives

$$(G^*(\mathbf{r}_B, \mathbf{r}_A) + G(\mathbf{r}_B, \mathbf{r}_A)) |S(\omega)|^2$$

$$= 2\omega \int \kappa_i(\mathbf{r}, \omega) |S(\omega)|^2 G(\mathbf{r}_A, \mathbf{r}) G^*(\mathbf{r}_B, \mathbf{r}) dV$$

$$= 2\omega \int \int \kappa_i(\mathbf{r}_1, \omega) \delta(\mathbf{r}_1 - \mathbf{r}_2)$$

$$\times |S(\omega)|^2 G(\mathbf{r}_A, \mathbf{r}_1) G^*(\mathbf{r}_B, \mathbf{r}_2) dV_1 dV_2$$

$$= 2\omega \int G(\mathbf{r}_A, \mathbf{r}_1) q(\mathbf{r}_1) dV_1 \left(\int G(\mathbf{r}_B, \mathbf{r}_2) q(\mathbf{r}_2) dV_2 \right)^*$$

$$= 2\omega p(\mathbf{r}_A) p^*(\mathbf{r}_B). \tag{13}$$

Expression (12) is used for the third identity, above, and expression (10) for the last one. The sum of the causal and acausal Green's function thus follows from correlating the pressure fields caused by the random volume sources:

$$G^*(\mathbf{r}_B, \mathbf{r}_A) + G(\mathbf{r}_B, \mathbf{r}_A) = \frac{2\omega}{|S(\omega)|^2} p(\mathbf{r}_A) p^*(\mathbf{r}_B). \tag{14}$$

As in seismic interferometry for loss-less media,[1,4,7] one needs to divide by the power spectrum of the excitation to remove the imprint of this excitation on the recorded pressure $p(\mathbf{r}_A)$ and $p(\mathbf{r}_B)$.

IV. WHEN THE SURFACE INTEGRAL IS NONZERO

In practical applications, none of the conditions C1–C3 might be satisfied. This is, in fact, the case in formulations of seismic interferometry where the Green's function is extracted by correlating pressure fields that are excited by uncorrelated sources at the surface that bounds the volume (e.g., Ref. 7). This section investigates the relative roles of the surface and volume integrals in expression (9). For simplicity, I use, following Wapenaar et al.,[7] that the surface is far from the region of interest and that $\nabla G(\mathbf{r}, \mathbf{r}_0) \cdot d\mathbf{S} = ikG(\mathbf{r}, \mathbf{r}_0) dS = (i\omega/c) G(\mathbf{r}, \mathbf{r}_0) dS$. Inserting this in Eq. (9) gives

$$G^*(\mathbf{r}_B, \mathbf{r}_A) + G(\mathbf{r}_B, \mathbf{r}_A) = I_V(\mathbf{r}_B, \mathbf{r}_A) + I_S(\mathbf{r}_B, \mathbf{r}_A), \tag{15}$$

where the volume integral $I_V(\mathbf{r}_B, \mathbf{r}_A)$ is given by

$$I_V(\mathbf{r}_B, \mathbf{r}_A) = 2\omega \int \kappa_i(\mathbf{r}, \omega) G(\mathbf{r}_A, \mathbf{r}) G^*(\mathbf{r}_B, \mathbf{r}) dV, \tag{16}$$

and the surface integral $I_S(\mathbf{r}_B, \mathbf{r}_A)$ by

$$I_S(\mathbf{r}_B, \mathbf{r}_A) = 2 \oint \frac{1}{\rho c} G(\mathbf{r}_A, \mathbf{r}) G^*(\mathbf{r}_B, \mathbf{r}) dS. \tag{17}$$

In many applications, the attenuation is weak ($\kappa_i \ll \kappa_r$), and one might think that the volume integral is small compared to the surface integral. This, however, is not the case. Because of the attenuation, the surface integral decreases exponentially with increasing surface area, and goes to zero while the volume integral is finite. According to expression (15) the sum of the volume integral and the surface integral is independent of the size of the volume. This implies that the volume integral and the surface integral are, in general, both needed for the extraction of the Green's function. The stationary phase analysis of Sec. V shows that, for the special case of a homogeneous medium, this is indeed the case.

V. STATIONARY PHASE ANALYSIS OF THE SURFACE INTEGRAL AND VOLUME INTEGRAL

To better understand the relative roles of the volume and surface integrals of expressions (16) and (17), the special case of a homogeneous, attenuating medium is analyzed in this section, and the volume and surface integrals are solved in the stationary phase approximation. For a homogeneous medium, Eqs. (1) and (2) can be combined to give

$$\nabla^2 p + \omega^2 \kappa \rho p = i\omega \rho q. \tag{18}$$

The wave number is therefore given by

$$k = \omega \sqrt{\kappa \rho}. \tag{19}$$

Weak attenuation is considered; in this case the wave number is to first order in κ_i / κ_r given by

$$k = \omega \sqrt{\kappa_r \rho} \left(1 + \frac{i\kappa_i}{2\kappa_r} \right). \tag{20}$$

The phase velocity thus is given by

$$c = \frac{\omega}{k_r} = 1/\sqrt{\kappa_r \rho}, \tag{21}$$

and the imaginary component of the wave number by

$$k_i = \omega \kappa_i \rho c/2. \tag{22}$$

The Green's function solution for expression (18) is equal to

$$G(R) = -i\omega \rho \exp(-k_i R) \frac{e^{ikR}}{4\pi R}. \tag{23}$$

The geometry for the stationary phase analysis is shown in Fig. 1. A coordinate system is used whose origin is at the midpoint of the receiver positions $\mathbf{r}_A$ and $\mathbf{r}_B$ and whose z axis points along the receiver line. The distance between these points is denoted by R; hence $\mathbf{r}_A = (0, 0, -R/2)$, and $\mathbf{r}_B = (0, 0, R/2)$. A volume that is bounded by a surface at distance L from the origin is considered. The stationary phase analysis follows the treatment of Ref. 5. The stationary phase point of the integrals in expressions (16) and (17) is located

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Roel Snieder: Extracting the Green's function and attenuation   2639
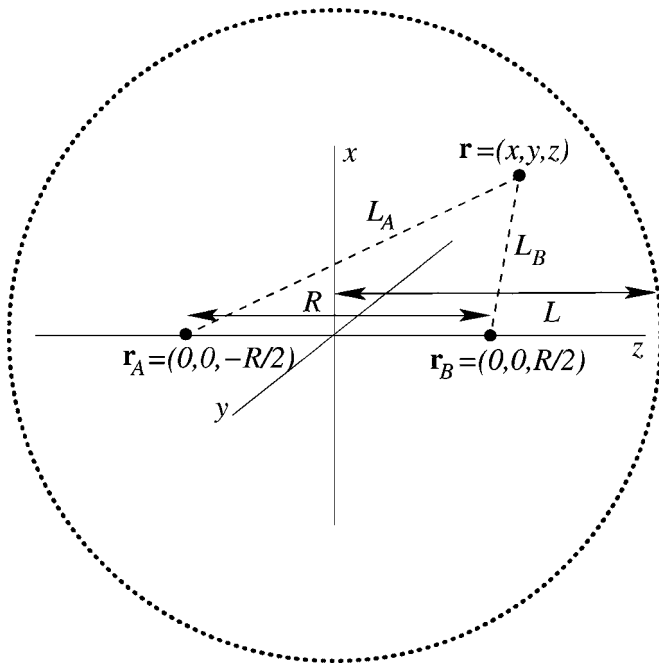
FIG. 1. Definition of geometric variables for the stationary phase evaluation of integrals $I_V(\mathbf{r}_B, \mathbf{r}_A)$ and $I_S(\mathbf{r}_B, \mathbf{r}_A)$. The volume is bounded by a sphere with radius L, as denoted by the dotted line.

FIG. 2. Two receivers (open squares) that are located between the acquisition surface and a reflector, and the stationary phase points SP_{dir} and SP_{refl} of the direct and reflected waves, respectively.

on the z axis ($x=y=0$). Following Ref. 5, the points to the right of $\mathbf{r}_B$ (for which $z>R/2$) give the causal Green's function $G(\mathbf{r}_B, \mathbf{r}_A)$, while the points to the left of $\mathbf{r}_A$ (for which $z<-R/2$) give the acausal Green's function $G^*(\mathbf{r}_B, \mathbf{r}_A)$. In the following only the contribution of integration points for which $z>R/2$ is treated; this gives only the causal Green's function. Because of this limitation, the corresponding surface and volume integrals are denoted with the superscript $(+)$.

Both the surface and volume integrals contain a double integration over the transverse x and y coordinates. As shown in the Appendix, the stationary phase approximation of the surface and volume integrals gives

$$I_V^{(+)}(\mathbf{r}_B, \mathbf{r}_A) = -i\omega\rho[\exp(-k_iR) - \exp(-2k_iL)]\frac{e^{ikR}}{4\pi R}, \quad (24)$$

and

$$I_S^{(+)}(\mathbf{r}_B, \mathbf{r}_A) = -i\omega\rho \exp(-2k_iL)\frac{e^{ikR}}{4\pi R}. \quad (25)$$

The sum of the surface and volume integrals indeed gives the causal Green's function:

$$I_S^{(+)}(\mathbf{r}_B, \mathbf{r}_A) + I_V^{(+)}(\mathbf{r}_B, \mathbf{r}_A) = -i\omega\rho \exp(-k_iR)\frac{e^{ikR}}{4\pi R}$$

$$= G(\mathbf{r}_B, \mathbf{r}_A). \quad (26)$$

Expressions (24) and (25) show that neither the volume integral nor the surface integral gives the Green's function, but that the sum does. Equation (16) suggests that for weak attenuation the volume integral can be ignored, because this integral is proportional to $\kappa_i \ll \kappa_r$. Expressions (24) and (25) show, however, for the special case of a homogeneous medium that as long as $k_iL=O(1)$, the volume integral and the

surface integral in general have comparable strength. (As shown in expression (A5), the volume integral $I_V^{(+)}$ is proportional to κ_i/k_i, which according to expression (22) has a finite value as $\kappa_i \to 0$.) The relative contribution of the surface integral and the volume integral is weighted by $\exp(-2k_iL)$. As the volume occupies all space ($L \to \infty$), the surface integral vanishes ($I_S^{(+)} \to 0$) and the volume integral is given by $I_V^{(+)}(\mathbf{r}_B, \mathbf{r}_A) = -i\omega\rho \exp(-k_iR)e^{ikR}/4\pi R = G(\mathbf{r}_B, \mathbf{r}_A)$. This is the special case treated in Sec. III because in this limit the surface integral vanishes because of the large distance L traversed by the attenuating waves that are correlated.

Equation (26) shows that in the frequency domain the Green's function can be retrieved from the cross correlation of waves excited by a combination of volume sources and surface sources. A similar result was obtained in the frequency domain in expression (10) of Ref. 5 where an infinite volume is needed. Expression (26) of Ref. 33 gives a time-domain formulation of the retrieval of the Green's function. In the latter studies sources in a homogeneous attenuating medium were integrated over an infinite volume. Because of the infinite integration region, the surface integral (25) did not contribute in those studies. The relative role of the surface integral and the volume integral is important because in some applications sources are present only on a finite surface (e.g., Ref. 14).

In this example, only the direct wave arrives, and ignoring the volume integral leads to an overall amplitude error. Next, the example of interferometry for both the direct wave and a reflected wave is considered. Sources are placed on the acquisition surface shown in Fig. 2. Both the direct wave and a reflected wave propagate to receivers indicated with open squares. The points SP_{dir} and SP_{refl} shown in Fig. 2 indicate the stationary phase source locations for the direct and reflected waves, respectively.[8]

The direct wave contains contributions $\exp(-k_iL_{dir})$ $\times \exp(-k_i(L_{dir}+R))$ from the attenuation at the stationary points. Following the stationary phase analysis of Ref. 8, and taking the attenuation terms into account gives a contribution of the surface integral to the direct wave that is given by

$$u_{\text{dir}} \propto e^{-2k_i L_{\text{dir}}} e^{-k_i R} \frac{e^{ikR}}{R}. \qquad (27)$$

The reflected waves have a contribution at the stationary phase point from the attenuation $\exp(-k_i L_{\text{refl}})\exp(-k_i(L_{\text{refl}}+R_1+R_2))$, the reflected wave obtained from the surface integral satisfies in the stationary phase approximation[8]

$$u_{\text{refl}} \propto e^{-2k_i L_{\text{refl}}} r e^{-k_i(R_1+R_2)} \frac{e^{ik(R_1+R_2)}}{R_1 + R_2}, \qquad (28)$$

where r is the reflection coefficient of the interface. Both the direct and reflected waves thus obtained have the correct phase and geometrical spreading, but both contain an amplitude term [$\exp(-2k_i L_{\text{dir}})$ and $\exp(-2k_i L_{\text{refl}})$, respectively] that is due to neglecting the volume integrals. Since these amplitude terms are different for the direct wave and the reflected waves, neglecting the contribution of the volume integrals disrupts the relative amplitude of the different arrivals.

It is interesting to compare this result with expressions (1) and (18) of Sabra *et al.*,[21] who show that for a homogeneous attenuating oceanic wave guide with source placed at a surface of constant depth that the phase of the different arrivals is correctly produced by the cross correlation, but the amplitude is not in the presence of attenuation. Expressions (27) and (28) presented here describe what happens when the sources are placed on a surface only. It is the absence of volume sources in Ref. 21 that leads to an incorrect estimate attenuation in the Green's function estimated from cross correlation.

VI. DISCUSSION

The derivation in this work shows that the Green's function of attenuating acoustic waves in a heterogeneous medium can be extracted by cross-correlating measurements of the pressure that is excited by random sources. As shown in Secs. III and IV, the Green's function can, however, be computed from the cross correlation when the random pressure field is excited by sources that are distributed throughout the volume, and that have a source strength that is proportional to the local dissipation rate (which is proportional to κ_i). Volume sources are also required for the extraction of the Green's function of the diffusion equation,[40] which is another example of a system that is not invariant for time-reversal.

The physical reason that the excitation must be proportional to the local dissipation rate is that the extraction of the Green's function is based on the equilibration of energy. This condition is necessary for the fluctuation-dissipation theorem, which relates the response of a dissipative system (the Green's function) to the fluctuations of that system around the equilibrium state.[41,42] Acoustic, dissipative, waves can be in equilibrium only when the excitation of the waves matches the local dissipation rate. If this were not the case, there would be a net energy flow, and the system would not be in equilibrium. The equilibrium of energy,[43] also referred to as equipartitioning, has been shown to be essential for the accurate reconstruction of the Green's function (e.g., Refs. 11, 30, and 44).

When none of the conditions C1–C3 of Sec. III is satisfied, the sum of the causal and acausal Green's function is, according to expression (15) given by the sum of the volume integral and a surface integral. The physical reason is that in equilibrium, the sources at the surface must be supplemented with sources within the volume that are proportional to the local dissipation rate if the system is to be in equilibrium. In some applications this condition can be realized. For example, Weaver and Lobkis[9] extract the Green's function from the wave field that is excited by thermal fluctuations throughout the volume of their sample. The need to have sources throughout the volume in addition to sources at the surface is impractical in applications where one seeks to extract the Green's function for two points in the interior by placing sources at the bounding surface only (e.g., Refs. 13 and 14).

Roux *et al.*[33] show that for a homogeneous infinite acoustic medium one needs to correct for a factor ω^{-1}. They note that this term is due to their assumed attenuation mechanism [$\text{Im}(c)=\text{constant}$]. In the formulation of this work, such a correction term is hidden in condition (12) which states that the power of the sources is proportional to the local attenuation rate. In this work, $\kappa_i(\mathbf{r}, \omega)$ can be an arbitrary function of position and frequency, but, as long as condition (12) is satisfied the Green's function can be extracted by cross correlation. In practical applications the source spectrum may not satisfy this condition. In that case there is no energy balance, and the Green's function is not correctly retrieved. This may be an important limitation in practical applications.

In practical situations, attenuation is present, and the contribution of the volume integral is often ignored, yet seismic interferometry seems to be able to retrieve the Green's function well (e.g., Refs. 13 and 14). For the special case of a homogeneous medium, the contribution of the surface integral to the Green's function is given by expression (25). This contribution has the correct phase and geometrical spreading [$\exp(ik_r R)/R$], but incorrect attenuation [$\exp(-2k_i L)$ instead of $\exp(-k_i R)$]. This suggests that when seismic interferometry for attenuating systems is used by summing over sources at the surface only, the correct phase and geometrical spreading are recovered, but that the attenuation is not. According to expressions (15)–(17) one needs for a general inhomogeneous attenuating medium both volume sources and surface sources for the extraction of the Green's function. It is known that multiple scattering by a boundary[45] or by internal scatterers[46] can compensate for a deficit of sources needed for focusing by time-reversal. This raises the unsolved question to what extent multiple scattering can compensate for the lack of volume sources.

APPENDIX: STATIONARY PHASE ANALYSIS OF THE INTEGRATION OVER THE TRANSVERSE COORDINATES

The integrals $I_V(\mathbf{r}_B, \mathbf{r}_A)$ and $I_S(\mathbf{r}_B, \mathbf{r}_A)$ of expressions (16) and (17) contain, in the geometry of Fig. 1, an integration over the x and y coordinates. This Appendix considers the contribution from integration points $z > R/2$. These points lead to the causal Green's function. The contribution from integration points $z < -R/2$ leads to the acausal Green's function, which can be obtained by complex conjugation of the results derived here.

For the Green's function of the homogeneous medium of expression (23),

$$G(\mathbf{r}_A, \mathbf{r})G^*(\mathbf{r}_B, \mathbf{r}) = \left(\frac{\rho\omega}{4\pi}\right)^2 e^{-k_i(L_A + L_B)} \frac{e^{ik_r(L_A + L_B)}}{L_A L_B}, \tag{A1}$$

where $L_{A,B} = |\mathbf{r} - \mathbf{r}_{A,B}|$, as shown in Fig. 1. The phase term $\exp(ik_r(L_A + L_B))$ of expression (A1) is oscillatory as a function of the transverse coordinates x and y. The phase is stationary along the z axis ($x = y = 0$). For fixed z, near the stationary phase point, the lengths L_A and L_B are, to second order in x and y, given by

$$L_A = \sqrt{x^2 + y^2 + (z + R/2)^2} \approx (z + R/2) + \frac{1}{2}\frac{x^2 + y^2}{(z + R/2)}, \tag{A2}$$

and

$$L_B = \sqrt{x^2 + y^2 + (z - R/2)^2} \approx (z - R/2) + \frac{1}{2}\frac{x^2 + y^2}{(z - R/2)}. \tag{A3}$$

These expressions are valid for integration points $z > R/2$. In the stationary phase approximation[47] of the integration of expression (A1) over x and y, these approximations for L_A and L_B are used in the phase term $\exp(ik_r(L_A + L_B))$. In the stationary phase approximation, the attenuation and geometrical spreading terms $\exp(-k_i(L_A + L_B))/L_A L_B$ are evaluated at the stationary phase point $x = y = 0$, where $L_{A,B} = z \pm R/2$. The integral of expression (A1) over the transverse coordinates is, in the stationary phase approximation, given by[47]

$$\int \int G(\mathbf{r}_A, \mathbf{r})G^*(\mathbf{r}_B, \mathbf{r})dxdy$$

$$= \left(\frac{\rho\omega}{4\pi}\right) \frac{e^{-2k_i z}}{z^2 - R^2/4} e^{ik_r R} \int \int \exp\left(-\frac{ik_r}{2}\left(\frac{R}{z^2 - R^2/4}\right)\right.$$

$$\left.\times (x^2 + y^2)\right)dxdy$$

$$= \left(\frac{\rho\omega}{4\pi}\right) \frac{e^{-2k_i z}}{z^2 - R^2/4} e^{ik_r R}\left(e^{-i\pi/4}\sqrt{\frac{2\pi(z^2 - R^2/4)}{k_r R}}\right)^2$$

$$= -\frac{i\rho^2\omega c}{8\pi} e^{-2k_i z}\frac{e^{ik_r R}}{R}. \tag{A4}$$

Inserting this result in Eq. (17), and setting $z = L$, gives expression (25) for the contribution of the surface $z = L$ to the surface integral $I_S(\mathbf{r}_B, \mathbf{r}_A)$. In order to obtain the contribution of the region $R/2 < z < L$ to the volume integral of expression (16), one needs to integrate Eq. (A4) over z:

$$I_V^{(+)}(\mathbf{r}_B, \mathbf{r}_A) = 2\omega\kappa_i \frac{-i\rho^2\omega c}{8\pi}\frac{e^{ik_r R}}{R}\int_{R/2}^{L} e^{-2k_i z}dz$$

$$= -\frac{i\rho^2\omega^2 c\kappa_i}{2k_i}\frac{e^{ik_r R}}{R}(e^{-k_i R} - e^{-2k_i L}). \tag{A5}$$

Using Eq. (22) to eliminate κ_i/k_i gives expression (24).

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

[2] A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from far-field correlations in an open scattering medium," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).

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

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

[5] 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).

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

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

[8] R. Snieder, K. Wapenaar, and K. Larner, "Spurious multiples in seismic interferometry of primaries," Geophysics **71**, SI111–SI124 (2006).

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

[10] R. Weaver and O. Lobkis, "On the emergence of the Green's function in the correlations of a diffuse field: Pulse-echo using thermal phonons," Ultrasonics **40**, 435–439 (2003).

[11] A. Malcolm, J. Scales, and B. A. van Tiggelen, "Extracting the Green's function from diffuse, equipartitioned waves," Phys. Rev. E **70**, 015601 (2004).

[12] E. Larose, G. Montaldo, A. Derode, and M. Campillo, "Passive imaging of localized reflectors and interfaces in open media," Appl. Phys. Lett. **88**, 104103 (2006).

[13] A. Bakulin and R. Calvert, "Virtual source: New method for imaging and 4D below complex overburden," Expanded Abstracts of the 2004 SEG-Meeting, pp. 2477–2480 (Society of Exploration Geophysicists, Tulsa, OK).

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

[15] M. Zhou, J. Jiang, Z. ad Yu, and G. T. Schuster, "Comparison between interferometric migration and reduced-time migration of common-depth-point data," Geophysics **71**, SI189–SI196 (2006).

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

[17] N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).

[18] K. G. Sabra, P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler, "Extracting time-domain Green's function estimates from ambient seismic noise," Geophys. Res. Lett. **32**, L03310 (2005).

[19] A. Paul, M. Campillo, L. Margerin, E. Larose, and A. Derode, "Empirical

synthesis of time-asymmetrical Green functions from the correlation of coda waves," J. Geophys. Res. **110**, B08302 (2005).

[20] P. Roux, W. A. Kuperman, and NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).

[21] K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-averaged ambient noise cross-correlation in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).

[22] K. G. Sabra, P. Roux, A. M. Thode, G. L. D'Spain, and W. S. Hodgkiss, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).

[23] R. Snieder and E. Şafak, "Extracting the building response using seismic interferometry; theory and application to the Millikan library in Pasadena, California," Bull. Seismol. Soc. Am. **96**, 586–598 (2006).

[24] R. Snieder, J. Sheiman, and R. Calvert, "Equivalence of the virtual source method and wavefield deconvolution in seismic interferometry," Phys. Rev. E **73**, 066620 (2006).

[25] J. E. Rickett and J. F. Claerbout, "Acoustic daylight imaging via spectral factorization; Helioseismology and reservoir monitoring," The Leading Edge **18**, 957–960 (1999).

[26] J. E. Rickett and J. F. Claerbout, "Calculation of the sun's acoustic impulse response by multidimensional spectral factorization," Sol. Phys. **192**, 203–210 (2000).

[27] J. E. Rickett and J. F. Claerbout, "Calculation of the acoustic solar impulse response by multidimensional spectral factorization," in *Helioseismic Diagnostics of Solar Convection and Activity*, edited by T. L. Duvall, J. W. Harvey, A. G. Kosovichev, and Z. Svestka (Kluwer Academic, Dordrecht, 2001).

[28] K. Wapenaar, D. Dragonv, and J. Robertsson, "Introduction to the supplement on seismic interferometry," Geophysics **71**, SI1–SI4 (2006).

[29] K. Wapenaar, E. Slob, and R. Snieder, "Unified Green's function retrieval by cross-correlation," Phys. Rev. Lett. **97**, 234301 (2006).

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

[31] O. A. Godin, "Recovering the acoustic Green's function from ambient noise cross correlation in an inhomogeneous medium," Phys. Rev. Lett. **97**, 054301 (2006).

[32] K. Wapenaar, "Nonreciprocal Green's function retrieval by cross correlation," J. Acoust. Soc. Am. **120**, EL7–EL13 (2006).

[33] P. Roux, K. G. Sabra, W. A. Kuperman, and A. Roux, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).

[34] K. Aki and P. G. Richards, *Quantitative Seismology*, 2nd ed. (University Science Books, Sausalito, 2002).

[35] J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975).

[36] A. T. de Hoop, "Time-domain reciprocity theorems for acoustic wave fields in fluids with relaxation," J. Acoust. Soc. Am. **84**, 1877–1882 (1988).

[37] A. T. de Hoop, *Handbook of Radiation and Scattering of Waves: Acoustic Waves in Fluids, Elastic Waves in Solids, Electromagnetic Waves* (Academic, San Diego, 1995).

[38] J. T. Fokkema and P. M. van den Berg, *Seismic Applications of Acoustic Reciprocity* (Elsevier, Amsterdam, 1993).

[39] E. Slob, D. Draganov, and K. Wapenaar, "GPR without a source," GPR 2006: 11th International Conference on Ground Penetrating Radar, Columbus, OH, 2006, Paper No. ANT. 6.

[40] R. Snieder, "Retrieving the Green's function of the diffusion equation from the response to a random forcing," Phys. Rev. E **74**, 046620 (2006).

[41] H. B. Callen and T. A. Welton, "Irreversibility and generalized noise," Phys. Rev. **83**, 34–40 (1951).

[42] R. Kubo, "The fluctuation-dissipation theorem," Rep. Prog. Phys. **29**, 255–284 (1966).

[43] R. L. Weaver, "On diffuse waves in solid media," J. Acoust. Soc. Am. **71**, 1608–1609 (1982).

[44] F. J. Sánchez-Sesma, J. A. Pérez-Ruiz, M. Campillo, and F. Luzón, "Elastodynamic 2D Green function retrieval from cross-correlation: Canonical inclusion problem," Geophys. Res. Lett. **33**, L13305 (2006).

[45] C. Draeger and M. Fink, "One-channel time reversal of elastic waves in a chaotic 2D-silicon cavity," Phys. Rev. Lett. **79**, 407–410 (1997).

[46] A. Derode, P. Roux, and M. Fink, "Robust acoustic time reversal with high-order multiple scattering," Phys. Rev. Lett. **75**, 4206–4209 (1995).

[47] N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Dover, New York, 1975).

Discomfort from sinusoidal oscillation in the roll and lateral axes at frequencies between 0.2 and 1.6 Hz

I. H. Wyllie and M. J. Griffin[a]

Human Factors Research Unit, Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, England

Discomfort caused by low frequency lateral and roll oscillations is often predicted from lateral acceleration in the plane of the seat, irrespective of whether it comes from horizontal motion or a component of gravity arising from roll. This study investigated discomfort from lateral and roll oscillation and whether acceleration in the plane of a seat predicts discomfort. Twelve subjects, sitting with and without backrest, used magnitude estimation to judge sinusoidal oscillations in the roll and lateral axes at ten frequencies between 0.2 and 1.6 Hz at magnitudes between 0.063 and 0.63 m s^{-2} root mean square. The rate of growth of vibration discomfort with increasing magnitude reduced with increasing frequency, so the frequency-dependence of discomfort varied with magnitude. Acceleration in the plane of the seat predicted discomfort from both lateral and roll oscillation at frequencies less than 0.4 Hz. At higher frequencies, acceleration produced by roll oscillation resulted in greater discomfort than the same acceleration produced by lateral oscillation. At frequencies greater than 0.4 Hz, a full height backrest increased discomfort with both lateral and roll oscillation. The prediction of discomfort caused by low frequency lateral and roll oscillation requires that both components are measured and assessed according to their separate effects.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2715654]

I. INTRODUCTION

The prediction of vibration discomfort in transport environments requires knowledge of the relationships between the physical environment (e.g., the magnitude, frequency, direction, and duration of vibration, and seating) and subjective reactions. The relationships are often represented by "weightings" for the effects of each factor. The relative importance of the influencing factors may be determined from systematic laboratory investigations designed to determine frequency weightings, axis weightings, etc. The accuracy of the procedures evolved for predicting vibration discomfort depends on the appropriateness of the weightings determined in such experiments.

In many travel environments (e.g., road vehicles, off-road vehicles, trains, boats, aircraft) there are substantial low frequency (i.e., 0.1–2.0 Hz) motions that may influence the comfort of passengers and operators. Although the contribution of low frequency vertical oscillation to discomfort has been studied (e.g., Shoenberger, 1975), human responses to low frequency fore-and-aft and lateral oscillation have been less investigated. A study over the frequency range 0.5–300 Hz suggested that the discomfort arising from exposure to acceleration in the horizontal axes was independent of both frequency and axis between 0.5 and 1.0 Hz (Miwa 1967). In the frequency range 1–30 Hz, Rao and Jones (1978) concluded that sensitivity to both lateral and fore-and-aft acceleration was greatest at 2 Hz, with decreasing sensitivity above and below this frequency. Corbridge and

Griffin (1986) investigated discomfort caused by lateral vibration at frequencies between 0.5 and 5 Hz and found that lateral acceleration caused most discomfort at frequencies between 1.6 and 2.0 Hz, with a gradual reduction in sensitivity at lower frequencies.

There has been very little investigation of the discomfort caused by oscillation in the rotational axes. Parsons and Griffin (1982) found that the discomfort caused by roll, pitch and yaw rotational acceleration of a seat decreased as the frequency of oscillation increased above 1 Hz, the lowest frequency studied. However they suggest caution in the interpretation of their results because the feet of subjects were supported on a stationary surface. Parsons and Griffin (1978) reported that with increasing distance of a seat from the centre of rotation, the frequency dependence of the equivalent comfort contours arising from 1 to 30 Hz rotational oscillation became similar to that for translational oscillation having a magnitude equivalent to the basal chord of the rotation—implying that when subjects are sufficiently far from the centre of rotation their discomfort is dominated by the translational motion caused by the rotation rather than by their rotation.

Current standards require that measurements of human exposure to whole-body vibration are made at the interfaces between the seat and the human body in the vertical and horizontal directions (e.g., BSI 1987; ISO 1997). Guidance is given for the measurement and evaluation of both translational and rotational oscillation. However, low frequency oscillation in the roll and pitch axes results in acceleration being measured in the lateral and fore-and-aft axes, respectively, due to the rotation of translational accelerometers relative to the gravity vector. At high frequencies of

---
[a]Electronic mail: M.J.Griffin@soton.ac.uk

vibration, where there are normally only small angles of rotation, the acceleration measured as a result of tilting through the gravitational vector is small. At low frequencies, where the angle of tilt can be large, the gravitational component may be significant. Consequently, the accelerations commonly measured by accelerometers orientated in the horizontal directions in vehicles are partly due to horizontal oscillation and partly due to the gravitational components arising from roll and pitch oscillation, but they are currently evaluated as though they are caused only by horizontal motion. It is unclear whether this method of evaluation provides a good prediction of discomfort.

The current standards also provide a simple and convenient means of combining the effects of vibration at the seat, the back, and the feet. The method assumes that any differences in phase between the motions at the seat, the back, and the feet can be neglected, although it is known that the effects of phase can be important at some frequencies (e.g., Jang and Griffin, 1999, 2000). At low frequencies, where there is little phase difference between motions at the seat and the feet, the methods in current standards are suspect since they imply that vibration at the feet will degrade discomfort, whereas support for the feet so that they move with the seat can be more comfortable than sitting with the feet either stationary or unsupported.

The objective of the present study was to determine how discomfort depends on the frequency of oscillation (at frequencies between 0.2 and 1.6 Hz), the direction of oscillation (lateral and roll) and seating condition (with and without a backrest). It was hypothesised that discomfort would vary similarly with roll and lateral oscillation when these motions had the same acceleration in the plane of the seat.

Two experiments are reported. A "between axes" experiment investigated the relative discomfort caused by roll and lateral oscillation at 0.2 Hz. A "within axis" experiment investigated the rate of growth in discomfort with increasing vibration magnitude, the variation in discomfort with the frequency of oscillation between 0.2 and 1.6 Hz in the roll and lateral directions, and the effect of a backrest. The within axis experiment also investigated the principal locations of discomfort in the body.

II. METHOD

Both experiments employed a 'within subjects' experimental design. Subjects were exposed to a reference stimulus followed by a test stimulus and asked to judge the discomfort of the test motion relative to the reference motion using the method of magnitude estimation (Stevens, 1975). The reference motions and all test motions were sinusoidal with durations of 30 s.

In the between axes experiment, the reference motion was lateral oscillation at a frequency of 0.2 Hz and a magnitude of 0.45 m s^{-2} root mean square (rms) with subjects seated on a flat rigid seat with a backrest and harness. The test motions were also at 0.2 Hz but with a magnitude chosen randomly from an array of magnitudes in either the lat-

eral or the roll axis and with or without a backrest and harness. This experiment was completed in one session with the order of presentation of conditions balanced.

In the within axis experiment, the reference motion was at a frequency of 0.5 Hz and a magnitude of 0.315 m s^{-2} rms. The test motions had frequencies between 0.2 and 1.6 Hz and were chosen randomly from an array of frequencies and magnitudes. There were four separate sessions, corresponding to motion in both axes (lateral or roll) and both postures (with or without a backrest with harness). The order of sessions was balanced and subjects attended only one session per day. For one magnitude of oscillation at each frequency, subjects indicated verbally the location in their body at which they felt the greatest discomfort. With lateral oscillation, they judged the location of discomfort with 0.4 m s^{-2} rms. With roll oscillation, they judged the location of discomfort with 0.4 ms^{-2} r.m.s. at frequencies between 0.2 and 0.5 Hz, and at magnitudes that decreased in inverse proportion to frequency at higher frequencies.

A. Stimuli in the between axes experiment

In the between axes experiment there were five 0.2 Hz stimuli with magnitudes in logarithmic series between 0.2 and 1.0 m s^{-2} rms. The simulator reproduced the test stimuli with unweighted acceleration distortions less than 8% in the lateral direction and less than 4% in roll. For the lateral motions, the cross-axis acceleration in the vertical direction (i.e., z axis) was less than 8% of the lateral (y axis) acceleration.

B. Stimuli in the within axis experiment

In the within axis experiment there were ten frequencies, one at each preferred one-third octave center frequency between 0.2 and 1.6 Hz. With lateral oscillation, there were 59 pairs of stimuli: six magnitudes in logarithmic series between 0.2 and 0.63 m s^{-2} rms at each frequency, except 0.2 Hz where five magnitudes were used in the range 0.2–0.5 m s^{-2} rms due to simulator displacement limitations.

With roll oscillation, there were 60 pairs of stimuli: six magnitudes at each frequency between 0.2 and 1.6 Hz. At frequencies between 0.2 and 0.5 Hz, the magnitudes expressed in terms of the equivalent lateral acceleration (i.e., the acceleration given by g sin θ, where θ is the angle of roll) were in a logarithmic series between 0.2 and 0.63 m s^{-2} rms. At frequencies greater than 0.5 Hz, the acceleration magnitudes decreased in inverse proportion to frequency such that at 1.6 Hz the magnitude varied between 0.063 and 0.2 m s^{-2} rms in a logarithmic series.

Roll oscillation was reproduced with unweighted acceleration distortion less than 8%. Lateral oscillation was reproduced with unweighted acceleration distortion less than 28%, and much less than 20% for the majority of stimuli. The frequency-weighted distortion of the lateral acceleration (weighted using W_d from BSI 1987) which should better reflect the perceptual impact of the distortion) was less than 20% at 0.63 and 1.0 Hz, and between 7 and 16% at the other eight frequencies.

Each of the four sessions lasted about 1 h and included at least one rest period.

C. Subjects

In both experiments there were 12 male subjects. All subjects were staff or students of the University and aged between 18 and 30 years. In the between axes experiment the mean subject mass was 74.5 kg (range 65–100 kg) and the mean stature 1.82 m (1.70–1.89 m). In the within axis experiment the mean mass was 75.6 kg (63.5–100 kg) and the mean stature 1.78 m (1.69–1.86 m).

D. Equipment

1. Seating

Rigid steel and aluminium alloy seats with flat horizontal wooden seat pans and removable vertical backrests were used in both experiments. The height of the seat pan was 420 mm above the simulator platform. When used, the height of the top of the backrest was adjusted to the sitting shoulder height of each subject—the backrest was adjustable in 30 mm increments within the 5th–95th percentile of British male adult sitting shoulder heights (Pheasant, 1996).

For the "harnessed" condition, subjects sat back in the seat in contact with the backrest and wore a four-point harness. The harness restrained subjects around the waist and at each shoulder. The harness was loosened before each test session and the subjects were instructed to tighten the harness to a "comfortably tight" setting, adjusting first the waist and then the shoulder restraints. The experimenter assisted the subject where necessary and ensured that the harness was symmetrically adjusted.

2. Equipment used in the between axes experiment

Oscillatory motion was produced by a simulator capable of 12 m of horizontal motion and ±10° of rotation located in the Human Factors Research Unit at the Institute of Sound and Vibration Research. Acceleration was measured using accelerometers operating on the principle of variable reluctance (Smiths type AV-L-692). Lateral oscillation was measured with an accelerometer mounted on a non-rotating portion of the chassis immediately below the seat. Roll oscillation was measured with a similar accelerometer located at the center of roll, which passed through the center of the seat surface.

3. Equipment used in the within axis experiment

Lateral oscillation was produced by a 1 m stroke horizontal hydraulic vibrator in the Human Factors Research Unit at the Institute of Sound and Vibration Research. The acceleration was measured using a capacitive accelerometer (Setra type 141A) mounted on the platform of the vibrator immediately below the seat.

To produce rotational oscillation, the same 1 m stroke horizontal vibrator was coupled via a crank to a rotational simulator. The axis of rotation passed through the centre of the seat surface and the acceleration arising from tilt through the gravitational vector (i.e., $g \sin \theta$) was measured using a capacitive accelerometer (Setra type 141A) orientated laterally and mounted at the center of rotation.

4. Signal generation and acquisition

The motion stimuli were generated and monitored using an HVLab data acquisition system and HVLab software (version 3.81). The drive signals were converted from digital to analog at 30 samples per second and low pass filtered at 10 Hz. Analog to digital conversion of the measured acceleration also took place at 30 samples per second after low pass filtering at 10 Hz.

E. Procedure

Prior to participating in the study, which was approved by the Human Experimentation Safety and Ethics Committee of the Institute of Sound and Vibration Research, subjects were screened using a list of medical contraindications (BSI 1973) and instructed in the method of magnitude estimation. Subjects then practiced judging the lengths of lines using the method of magnitude estimation before an experiment commenced.

Subjects were asked to sit in a comfortable upright posture, with their feet shoulder-width apart, and to use the method of magnitude estimation to express the vibration discomfort caused by the test motions relative to the discomfort caused by the reference motion, ignoring any audible noise. The subjects assigned a number that represented the discomfort of the test motion relative to the discomfort of the reference motion, assuming the discomfort caused by the reference motion corresponded to "100." Subjects were permitted to use any positive number for their estimate of discomfort.

During the between axes experiment, subjects sat in a cabin that prevented them seeing their movement relative to the laboratory. The experimenter was able to observe subjects at all times via a closed circuit television system. During the within axis experiment, subjects wore a blindfold. In both experiments, subjects wore headphones through which noise at 80 dB (A) was delivered to mask the operating noise of the simulator. The headphones also facilitated communication between the experimenter and the subject.

III. RESULTS

A. Rate of growth of discomfort

The magnitudes of the physical stimuli, φ (accelerations in the plane of the floor), were related to the magnitudes of the sensations, ψ (magnitude estimates of discomfort), using Stevens' power law (Stevens, 1975):

$$\psi = k\varphi^n. \tag{1}$$

The rates of growth in discomfort, n, were determined for each subject by regression between the logarithm of the vibration magnitude, φ, and the logarithm of the magnitude estimates, ψ:

$$\log_{10} \psi = \log_{10} k + n \log_{10} \varphi \tag{2}$$

Median rates of growth at each frequency were calculated from the individual slopes, n, and individual intercepts, k (Fig. 1). The data from one subject in the within axis experiment whose judgements indicated lack of comprehension of the method of magnitude estimation were discarded.
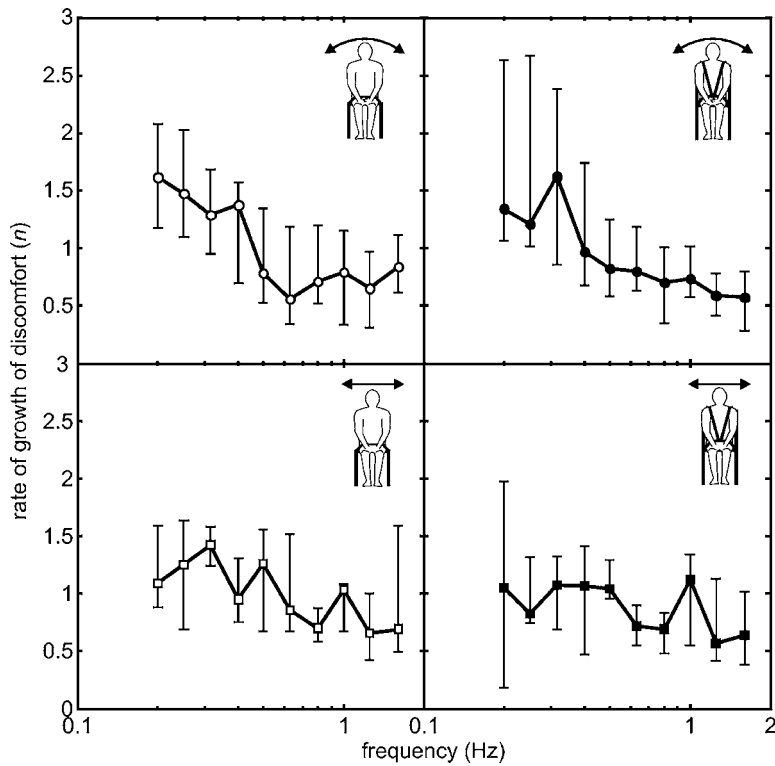
FIG. 1. Median rates of growth of discomfort, n, for roll and lateral oscillation with and without a backrest. Upper and lower error bars show the 75th and 25th percentile rates of growth of discomfort, respectively.

With roll oscillation, the rate of growth of vibration discomfort with increasing vibration magnitude varied significantly with frequency, both when subjects were seated on a flat rigid seat ($p<0.01$; Friedman) and when they were seated on a rigid seat with a backrest ($p<0.05$; Friedman). Analysis of the trend showed that the median rate of growth of discomfort was negatively correlated with increasing frequency ($p<0.05$; Spearman).

With lateral oscillation and no backrest, the rate of growth of discomfort varied significantly with frequency ($p<0.01$; Friedman) and analysis of the trend showed that the median rate of growth of discomfort was negatively correlated with increasing frequency ($p<0.01$; Spearman). With the backrest, the rate of growth of discomfort did not vary significantly with frequency ($p>0.05$; Friedman).

The presence of a backrest had no statistically significant effect on the rate of growth of discomfort with either roll or lateral oscillation at any frequency ($p>0.05$ Wilcoxon).

With no backrest, the median rate of growth of discomfort was lower for roll than for lateral oscillation at 0.315, 0.5, 0.63, 1.0, and 1.2 Hz, but the difference was only significant at 0.63 Hz ($p<0.05$ Wilcoxon). With the backrest, the median rate of growth of discomfort was greater for roll than for lateral oscillation at 0.2, 0.25, 0.315, 0.63, 0.8, and 1.25 Hz but the difference was only significant at 0.25 and 0.315 Hz ($p<0.05$, Wilcoxon). Adjustment according to the Bonferroni criterion suggests that these differences with and without a backrest might have arisen by chance.

The rates of growth in discomfort determined at 0.2 Hz during the between axes experiment and the within axes experiment were not significantly different for either roll or lateral oscillation ($p>0.05$, Mann-Whitney U), except for roll oscillation without backrest where the rate of growth of vibration discomfort was greater in the single frequency between axes experiment ($n=1.09$) than during the multiple frequency 'within axes' experiment ($n=1.54$) ($p=0.012$ Mann-Whitney U).

B. Relative discomfort between lateral and roll oscillation

1. The relationship at 0.2 Hz

The magnitudes of 0.2 Hz oscillation required in both axes and with both seating conditions to elicit a discomfort judgment of 100 (i.e, discomfort equivalent to the 0.2 Hz sinusoidal reference motion at 0.45 m s^{-2} rms in the lateral axis on the rigid seat with backrest) differed significantly ($p<0.05$ Friedman). In order of increasing sensitivity, the equivalent magnitudes were: lateral on a flat rigid seat (0.37 m s^{-2} rms), lateral on a rigid seat with backrest (0.45 m s^{-2} rms), roll on a rigid seat with backrest (0.47 m s^{-2} rms), and roll on a flat rigid seat (0.54 m s^{-2} rms).

2. Comparing results of the within axis experiment across axes

In the between axis experiment it was found that the reference motion exerted an effect on discomfort that varied between axes and between seating conditions. To adjust the level of the contours obtained in the within axis experiment to allow for the differing sensitivity found at 0.2 Hz in the between axis experiment, a correction factor, C_{Si}, was calculated for each of the four conditions:

$$C_{Si} = \varphi_{Si}/\varphi_{SR} \tag{3}$$

where φ_{Si} is the magnitude of a 0.2 Hz motion in the between axis experiment that gave discomfort equivalent to

0.2 Hz lateral oscillation at 0.45 m s^{-2} rms on a rigid seat with backrest, and φ_{SR} is the magnitude of a 0.2 Hz test motion in the within axis experiment that corresponded to a magnitude estimate of 100 when using a 0.5 Hz 0.315 m s^{-2} rms sinusoidal reference motion in the same axis and seating condition as the test stimulus.

The correction factors were calculated from the medians of the individual slopes and intercepts. Stevens (1975) found that the n value was independent of the reference condition, so it was assumed that the reference motion had a constant effect at all frequencies and the same 0.2 Hz correction factor was applied at all frequencies.

Five subjects participated in both the within axis and the between axis experiments. Their k and n values were used to calculate individual correction factors. For each of the four conditions, the median correction factor from the 12 subjects lay within the range of the five individual subject correction factors.

To allow statistical comparisons across axes (see the following), the individual contours obtained in the within axis experiment were adjusted by the median correction factors obtained in the between axis experiment.

C. Effect of frequency on discomfort caused by lateral and roll oscillation

Equivalent comfort contours were determined within each axis for each subject by calculating the vibration acceleration, φ, corresponding to each of five subjective magnitudes, ψ, (i.e., 63, 80, 100, 125, and 160) where 100 corresponds to the discomfort produced by 0.315 m s^{-2} rms at 0.5 Hz in that axis, for each frequency (from 0.2 to 1.6 Hz) using Eq. (2). Five median equivalent comfort contours were then generated from the medians of the contours of individual subjects. The contour equivalent to a perceptual magnitude of 100 was then adjusted according to differences between axes as stated in Eq. (3). The adjusted contours for a subjective magnitude of 100 were highly dependent on vibration frequency ($p<0.01$, Friedman; Fig. 2).

With roll oscillation, the level of the 100 comfort contour declines at almost 12 dB per octave between 0.2 and 1.6 Hz when using a backrest and about 6 dB per octave when not using a backrest. With lateral oscillation, the level of the 100 comfort contour declines at very approximately 3 dB per octave between 0.2 and 1.6 Hz when not using a backrest, and about 6 dB per octave when using a backrest. Without a backrest at frequencies greater than 0.5 Hz, and with backrest at frequencies greater than 0.8 Hz, the comfort contours corresponding to a subjective magnitude of 100 were significantly lower for roll oscillation than for lateral oscillation—indicating greater sensitivity to roll oscillation than to lateral oscillation ($p<0.05$ Wilcoxon).

D. Effect of backrest

With roll oscillation, at frequencies less than about 0.5 Hz, the backrest had little effect on the level of the 100 comfort contour, although the median contour suggests slightly less discomfort without the backrest. At frequencies greater than 0.63 Hz, subjects were significantly more sensi-
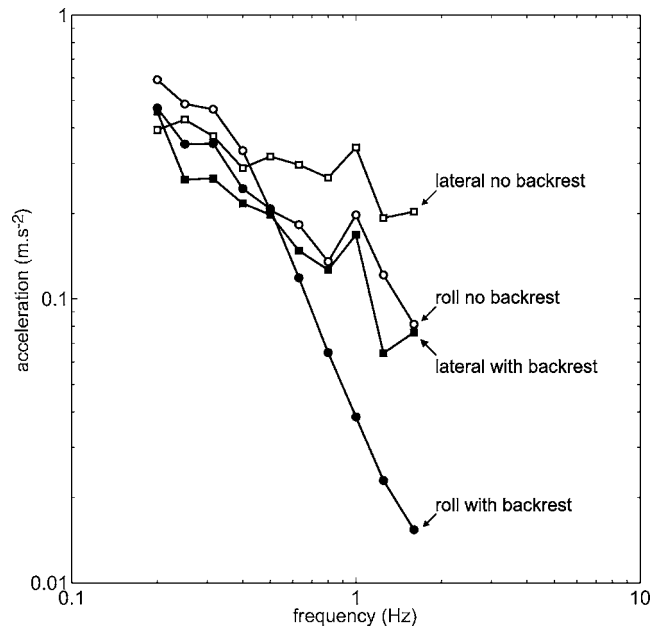
FIG. 2. Adjusted roll and lateral contours, each producing discomfort equivalent to that arising from exposure to sinusoidal lateral oscillation on a rigid seat with a backrest at 0.2 Hz, 0.45 m s^{-2} rms.

tive to the motion when the backrest was present ($p<0.05$ Wilcoxon). With lateral oscillation at 0.315 Hz, and at all frequencies greater than 0.4 Hz, subjects experienced greater discomfort when the backrest was present ($p<0.05$, Wilcoxon).

E. Effect of magnitude on the frequency dependence of equivalent comfort contours

The effect of motion magnitude on the level of the comfort contours equivalent to subjective judgements between 63 and 160 arising from exposure to motion in the roll and lateral axes is shown in Fig. 3. These median contours were calculated from individual subject contours produced using Eq. (2). The reference condition for these contours is 0.5 Hz, 0.315 m s^{-2} rms sinusoidal motion in the same axis as the test motion, so the levels of these contours in one axis should not be compared with those in another axis.

Within the range of stimuli employed in this study, the shapes of the equivalent comfort contours are not greatly affected by the magnitude of the stimuli, although it can be seen that the spread between contours is less where the n value is greatest at low frequencies.

F. Location of discomfort

There were no clear variations in the locations of discomfort with frequency. However there was a higher incidence of subjects reporting discomfort at the head, neck, or shoulders while seated with a backrest than without a backrest (Fig. 4). This trend reached significance at 1.0 Hz during exposure to roll oscillation ($p<0.01$, McNemar) and at 0.315 Hz during exposure to lateral oscillation ($p<0.05$, McNemar).
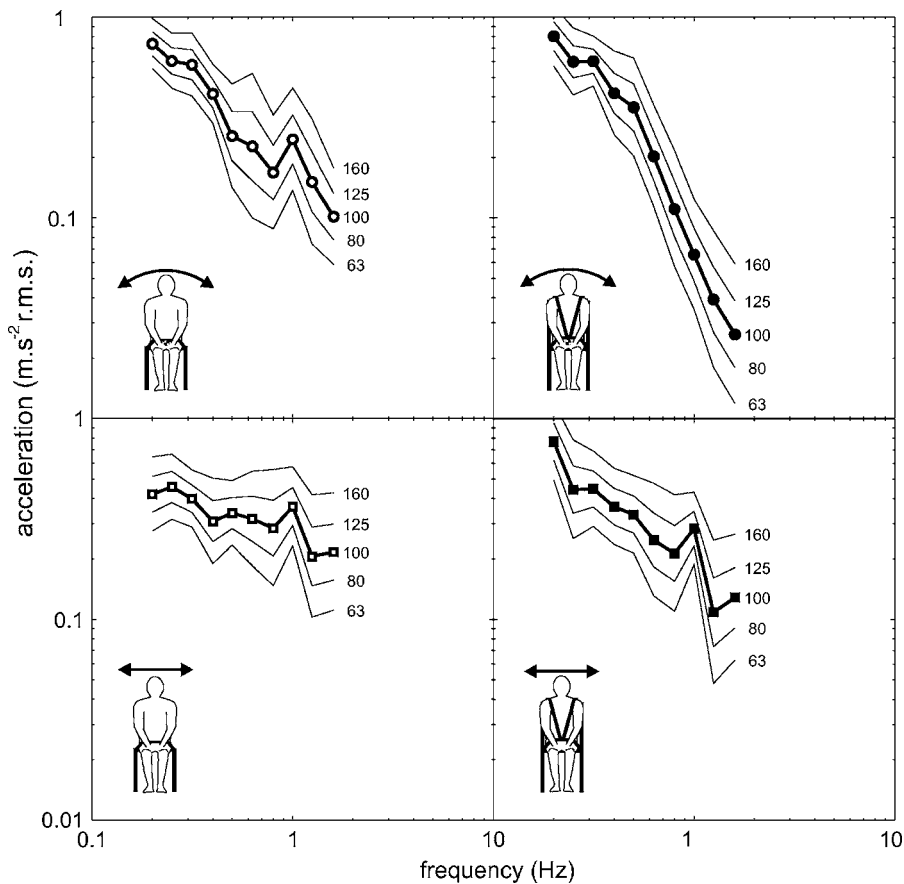
FIG. 3. The effect of frequency and magnitude on the level of the equivalent comfort contours arising from exposure to oscillation in the roll and lateral axes on both a flat rigid seat and a rigid seat with a backrest. The 100 contour (bold) represents discomfort equivalent to that caused by 0.5 Hz 0.315 m s^{-2} rms oscillation in the same axis.

IV. DISCUSSION

A. Rate of growth of discomfort caused by lateral and roll oscillation

From a study of low frequency vertical oscillation at frequencies between 0.2 and 1.6 Hz, Shoenberger (1975) reported rates of growth of vibration discomfort between 1.34 and 1.47. In the present study with lateral oscillation, the rates of growth of discomfort are similar, despite the differing axis and seating condition.

A high n value is associated with a greater increment in the subjective magnitude, ψ for a unit increase in the physical magnitude of the stimulus, φ. The results show that a unit increment in the magnitude of oscillation was associated with a greater increment in discomfort at low frequencies than at high frequencies. This means that contours of equivalent discomfort change shape with the magnitude of oscillation, as seen in Fig. 3. Over a wider range of magnitudes, the change of shape would be greater and it would be seen that a frequency weighting appropriate for low magnitudes would be inappropriate for high magnitudes.

During exposure to vibration at frequencies similar to, and greater than, those studied here, the presence of a backrest changes biodynamic responses of the body, including the lateral apparent mass (Fairley and Griffin, 1990), roll apparent mass (Gunston, 2004), the transmission of lateral vibration from the seat to the head (Paddan and Griffin, 1988) and
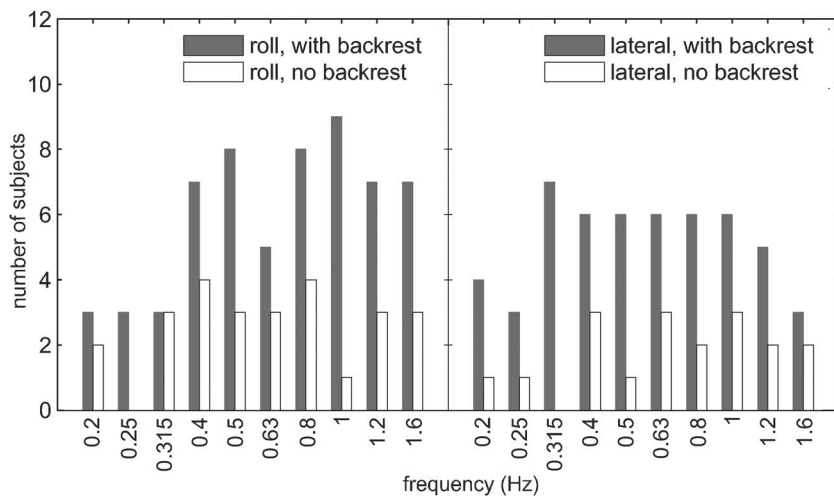
FIG. 4. Effect of backrest and frequency of oscillation on the incidence of the dominant location of discomfort being at the head, neck or shoulders in the 12 subjects.

the transmission of roll vibration from the seat to the head (Paddan and Griffin, 1994). Notwithstanding these biodynamic changes, the rate of growth of discomfort was not found to be affected by the backrest in either axis.

The rates of growth in discomfort with vibration magnitude (i.e., the value of n in Stevens' power law) showed a high level of inter-subject variability. At the lowest frequencies, subjects were able to exercise voluntary control of the movement of their upper bodies in response to motion, whereas at the higher frequencies the response may have been more dictated by involuntary muscular activity. The experimental design presented stimuli in a random order of frequency and magnitude, so subjects experienced motions of the same frequency at widely separated points in time. They may have chosen to respond differently to stimuli of the same frequency at different times, although no evidence of a systematic effect could be found in the data. Such variability in their chosen response would reduce correlations between the magnitudes of the motions and the magnitude estimates of discomfort made by subjects.

B. Relative discomfort between lateral and roll oscillation at 0.2 Hz

The rate of growth of discomfort at 0.2 Hz during exposure in the between axes and the within axis experiments showed little difference other than a marginally significant difference with roll oscillation. The difference might have arisen from the different visual conditions—subjects were exposed to motion in an enclosed cabin in the between axes experiment but blindfolded in the within axis experiment. Allowing subjects to see their motions may have altered their voluntary postural responses to the motions.

The interquartile range of n values in the within axis experiments tended to be greater than in the between axis experiments. The stimuli were randomized by frequency and magnitude in the within axis experiment and so this increased variability is consistent with subjects choosing to respond differently on the different occasions they were exposed to the same frequency.

The adjustments required for the four sets of comfort contours so that they represent the same degree of discomfort with both axes and both postures were not large. There is limited knowledge of what constitutes a just noticeable difference at the frequencies of this study. With 5 Hz vertical vibration of seated subjects, Morioka and Griffin (2000) and Forta (2005) have reported difference thresholds of 12.3% and 13.4% respectively. If such changes are required to detect differences in discomfort at 0.2 Hz, it can be argued that the contours obtained in the within axis experiment can be used without the adjustment calculated using the between axis experiment—although this would not have been known without the between axis experiment. The differing frequency-dependence between axes and postures (as evident in Fig. 2) means that the adjustment would have been large if a reference frequency much greater than 0.2 Hz had been employed in the between axes experiment.

C. Effect of frequency on discomfort caused by lateral and roll oscillation

At frequencies less than 0.4 Hz, the levels of the 100 comfort contours for lateral and roll oscillation are not the same, but they are very roughly similar (Fig. 2), indicating that a broadly similar degree of discomfort was caused by acceleration in the plane of the seat irrespective of whether the acceleration arose from lateral oscillation or roll through the gravitational vector. At frequencies greater than about 0.4 Hz, the differences increase and roll oscillation causes greater discomfort than lateral oscillation, with the difference increasing with increasing frequency of oscillation.

Caution over the similarity in response at frequencies less than about 0.4 Hz may be appropriate. In this frequency range, subjects can adopt different strategies and could choose to allow their bodies to move differently with lateral and roll oscillation. It is not yet clear which strategy gives least comfort and persons exposed to these motions may not always find the most comfortable strategy. Furthermore, although similar discomfort was found with the 30-second stimuli used here, differences in muscular activity in response to the two motions may result in other differences in discomfort with longer durations of exposure.

The finding that lateral acceleration in the plane of the seat does not predict discomfort at frequencies greater than 0.4 Hz means that the prediction of discomfort in vehicles where there is significant oscillation at frequencies greater than 0.4 Hz requires knowledge of whether the measured acceleration is caused by translation or caused by rotation through the gravitational vector. Many vehicles have suspensions causing roll oscillation at frequencies in the range 0.4–1.6 Hz when excited by undulations of the terrain or track or by changes of direction. The motions are often evaluated considering only the acceleration indicated by accelerometers orientated in the lateral axis of the vehicle, disregarding whether the measured acceleration arises from lateral acceleration or the component of gravity (i.e., g sin θ) caused by roll. The results of this study show that, in the frequency range 0.4–1.6 Hz, such acceleration is less uncomfortable when it is caused by lateral acceleration than when it is caused by roll.

D. Effect of backrest

With roll oscillation, the backrest increased discomfort at frequencies greater than about 0.5 Hz (Fig. 2). With constant acceleration in the plane of the seat (i.e., constant g sin θ, due to a constant angle of oscillation), the lateral acceleration at the top of the backrest increases in proportion to the square of the frequency of oscillation, so causing increased discomfort with increasing frequency. Without a backrest, there appeared to be less motion of the upper bodies and heads of subjects.

With low frequency lateral oscillation and no backrest, the motion of the upper body could be seen to lag behind the motion of the platform, suggesting there was a resonance with amplification of body motion and greater discomfort, or voluntary and involuntary back muscle activity producing sensations described as "discomfort." However, this varied
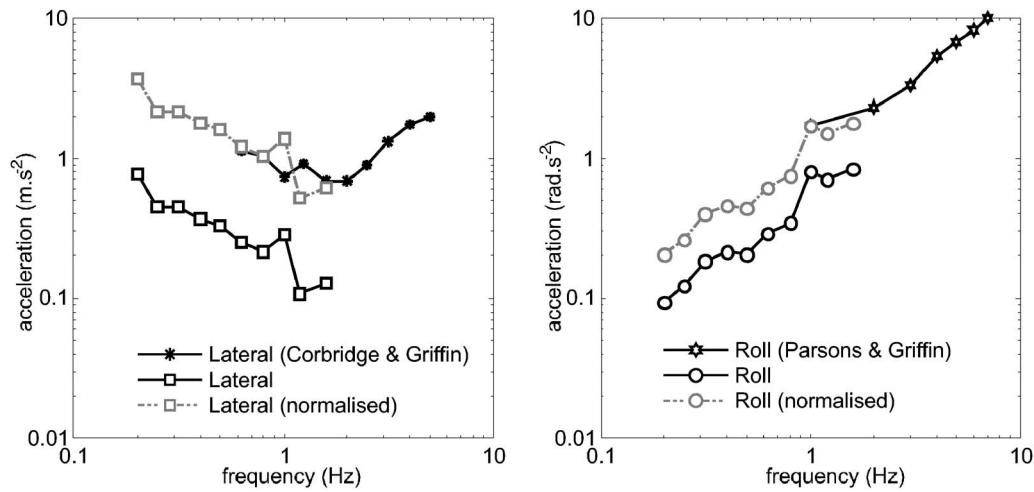
FIG. 5. Comparison of equivalent comfort contours without backrest with previous studies of lateral oscillation (Corbridge and Griffin, 1986) and roll oscillation (Parsons and Griffin, 1982).

between subjects. With a backrest and harness, the upper body was held and moved with the backrest, reducing any amplification due to resonance and reducing the opportunity of useful back muscle activity. Although the contours in Fig. 2 tend to suggest slightly greater discomfort at low frequencies when the backrest was present, this may depend on how a person chooses to "ride the motion" when there is no backrest.

The present study used two apparently extreme backrest conditions: no backrest and a full height backrest with harness. These two extremes may not represent extremes of the effects of backrests on the discomfort caused by low frequency oscillation. For example, the detrimental effects of a backrest with roll oscillation were partially caused by the lateral acceleration at the top of the backrest, so discomfort will be reduced with a lower backrest. Any beneficial effects of a backrest at low frequencies may therefore be extended towards higher frequencies as the height of a backrest is reduced.

E. Location of discomfort

At higher frequencies than studied here, the location of discomfort is strongly associated with the frequency of vibration (e.g., Griffin and Whitham, 1978). Jang and Griffin (2000) found that variations in the relative motion between the seat and the feet altered the location of discomfort at frequencies close to those used in the present study. It seems probable that at frequencies less than 1.6 Hz, oscillation produces discomfort which is not well localized, and that measures of static comfort may become increasingly important in predicting responses to the environment.

With both lateral and roll oscillation, the presence of a backrest increased the incidence of the dominant location of discomfort being identified at the head-neck interface. It seems that this was because the backrest prevented the torso moving so as to reduce the acceleration reaching the head and neck. Any increased muscular activity in the neck when using a backrest, and any increased muscular activity in the back when not using a backrest, may have different time-dependent characteristics. These time dependencies may

mean that comfort contours obtained during short exposures may not give a good indication of discomfort experienced during long exposures.

F. Comparison with previous research

Miwa (1967) reported the frequency dependence of discomfort caused by lateral vibration on a rigid seat with backrest at frequencies between 0.5 and 300 Hz. At supra threshold levels, sensitivity was independent of frequency at frequencies less than 2 Hz. The results of the present study show discomfort decreasing with decreasing frequency below 1.6 Hz. The difference between these results and those of Miwa (1967) might be due to the short exposures used by Miwa—at 0.5 Hz his subjects were exposed to only 3 cycles of oscillation.

With a minimum frequency of 1.0 Hz and similar magnitudes to those studied here, Rao and Jones (1978) described decreasing sensitivity as the frequency decreased from 2 to 1 Hz when subjects were exposed to oscillation on a seat with no backrest, consistent with the present findings.

Corbridge and Griffin (1986) investigated the discomfort caused by lateral oscillation at frequencies greater than 0.5 Hz using a rigid seat with a backrest. Although the magnitudes of the stimuli were greater than those used in the present study, the overall trends appear similar. When the results of this work are normalised with respect to those of Corbridge and Griffin (1986) at 0.8 Hz (the lowest frequency for which they had a complete dataset), the frequency-dependence for discomfort is similar to the present study (Fig. 5).

Parsons and Griffin (1982) investigated the discomfort caused by roll vibration at frequencies between 1 and 30 Hz with subjects seated on a flat rigid seat with no backrest. They found that sensitivity to roll acceleration (rad s^{-2}) increased with decreasing frequency. When the results of the present study are expressed in terms of angular acceleration, an overall trend for increased sensitivity to roll acceleration at lower frequencies continues to 0.2 Hz, although with some increase in sensitivity between 0.5 and 1.0 Hz (Fig. 5).
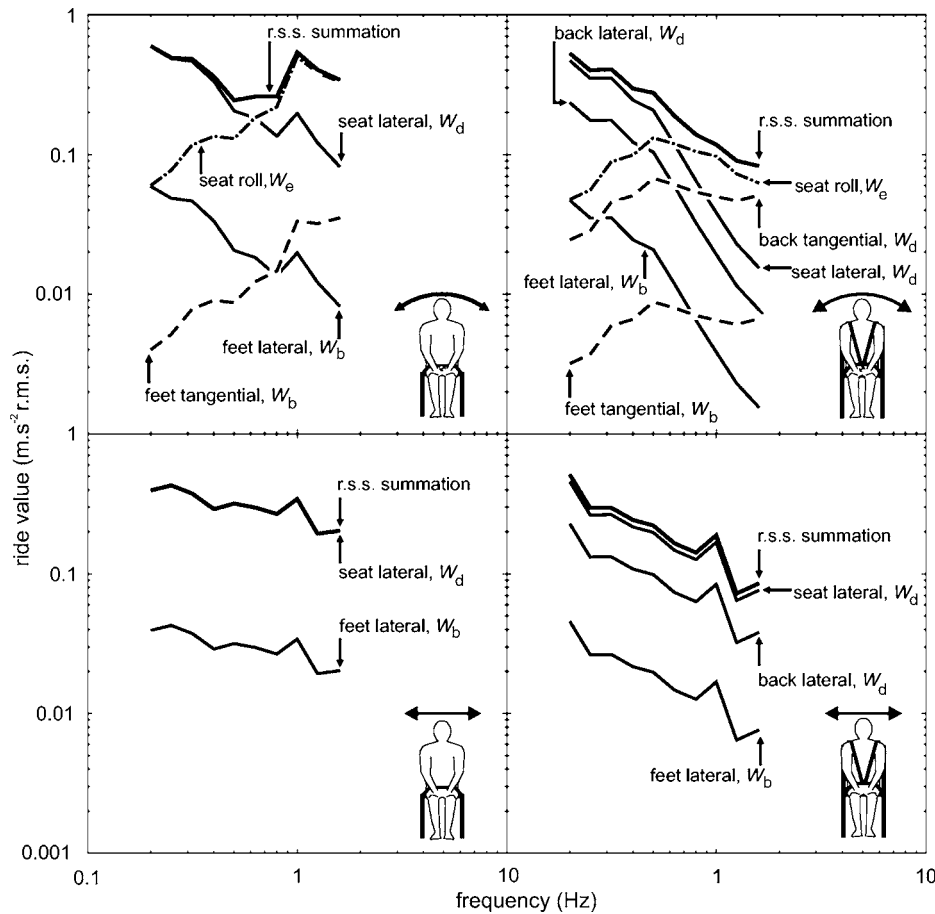
FIG. 6. Frequency-weighted accelerations corresponding to the adjusted contours generated from the median individual k and n values. [Values calculated using BS 6841 (BSI 1987) asymptotic frequency weightings extrapolated below 0.5 Hz without band-pass filtering. Components of motion shown together with root-sums-of-squares of weighted values in each axis and seating condition.

G. Prediction of discomfort caused by roll oscillation

According to current standards, the discomfort caused by mechanical oscillation in the frequency range 0.5–80 Hz can be predicted using frequency weightings for translational and rotational vibration at the seat and from translational vibration at the feet and the backrest (BSI 1987; Griffin, 1990). The method means that with combined roll and lateral oscillation, seven components can theoretically contribute to discomfort:

(i)   horizontal acceleration in the plane of the seat (weighted by frequency weighting W_d);

(ii)   translational acceleration in the plane of the seat due to roll–i.e. g . sin θ (weighted by frequency weighting W_d);

(iii)   roll acceleration in the plane of the seat (rad s^{-2}) (weighted by frequency weighting W_e, with a multiplying factor of 0.63);

(iv)   lateral acceleration at the backrest (weighted by frequency weighting W_d, with a multiplying factor of 0.5);

(v)   translational acceleration at the backrest in the plane of the seat due to roll—i.e., g . sin θ (weighted by frequency weighting W_d, with a multiplying factor of 0.5);

(vi)   lateral acceleration at the feet (weighted by frequency weighting W_b, with a multiplying factor of 0.25); and

(vii) translational acceleration at the feet in the plane of the seat due to roll—i.e., g sin θ (weighted by frequency weighting W_b, with a multiplying factor of 0.25);

In practice, many of these components are sufficiently small to be neglected. For example, with high frequencies of vibration the angle of roll is small, so g . sin θ is small and the contributions of components (ii), (v), and (vii) are insignificant.

Hitherto, it has been assumed that either the components due to g . sin θ are small (as with high frequency vibration) or that the effects are adequately reflected in the combination of this component with the lateral acceleration measured by nominally horizontal accelerometers. The results of the present study allow this assumption to be tested, but only provide definitive information about the relative discomfort of roll and lateral motion when the centre of roll is in the plane of the seat surface. This is not the case in some practical situations.

The four equivalent comfort contours shown in Fig. 2, correspond to conditions in which there was the same discomfort at all frequencies—hence the root-sums-of-squares (i.e., rss) summation of weighted components for each of these four conditions should be the same at all frequencies if the standardised weighting method in BS 6841 (BSI 1987) is appropriate – at least at frequencies greater than 0.5 Hz where the method is intended to be used. The contours in Fig. 6 were constructed from the four equivalent comfort
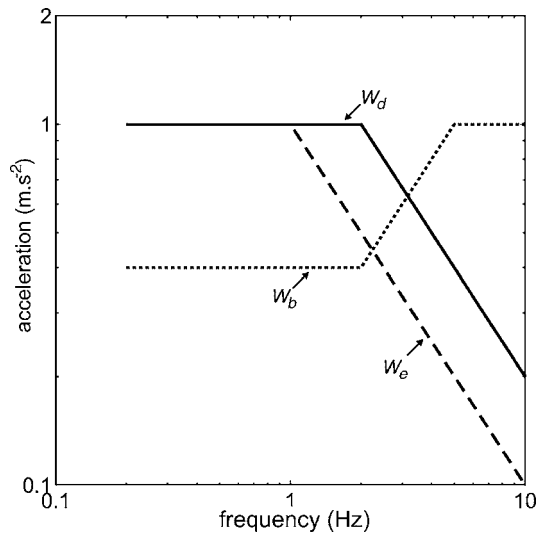
FIG. 7. Asymptotic frequency weightings according to BS 6841 (BSI 1987) extrapolated below 0.5 Hz without band-pass filtering.

contours in Fig. 2 (i.e. the motions that produced discomfort equivalent to 0.2 Hz 0.45 m s^{-2} rms lateral oscillation with a rigid seat and backrest) after applying the standardised frequency weightings without the high-pass filter at 0.4 Hz (equivalent to the simple asymptotic weightings shown in Fig. 7) with the multiplying factors as specified in BS 6841 (BSI 1987). The seat lateral acceleration contour in each figure was obtained by weighting the measured contour by W_d. The foot lateral acceleration contours were calculated from the foot lateral acceleration weighted by W_b with a multiplying factor of 0.25. Any roll acceleration at the seat was weighted by W_e with a multiplying factor of 0.63. Where the back was in contact with the backrest, the back lateral acceleration contour was weighted by W_d with a multiplying factor of 0.5, and the tangential acceleration at the top of the backrest due to roll was weighted by W_d with a multiplying factor of 0.5. The tangential acceleration at the feet due to roll was weighted by W_b with a multiplying factor of 0.25. The rss of the components present (two components with lateral oscillation without a backrest, three with lateral oscillation with backrest, four with roll oscillation without a backrest, and six with roll oscillation with a backrest) were then calculated.

The prediction method in BS 6841 (BSI 1987) allows for the assessment of discomfort caused by complex motions containing both rotational and translational components using weightings that assume the rotational components are measured (rad s^{-2}) and the translational components are measured (m s^{-2}). After frequency weighting according to the weightings and multiplying factors in the standard, components in any axis (translational or rotational) can be compared and combined such that the overall ride value is appropriately influenced by each component according to its importance.

For lateral oscillation without a backrest, the equivalent comfort contour (i.e., the rss of the weighted lateral seat and lateral floor accelerations) corresponds to a similar overall ride value at all frequencies from 0.5 to 1.6 Hz, as expected. The slight decrease with increasing frequency suggests that

sensitivity to lateral acceleration is slightly less than predicted at the lower frequencies. The main contributor to the overall ride value is the lateral oscillation at the seat, with the motion at the feet contributing little to the discomfort, according to the standardized method. As predicted, the contour is at about the same level as the reference motion (i.e., 0.2 Hz 0.45 m s^{-2} rms lateral oscillation with a rigid seat and backrest). At frequencies less than about 0.5 Hz, the contour remains at a similar level (whereas when the band-limiting filters at 0.4 Hz are used the contour drops sharply)—suggesting that constant sensitivity to lateral acceleration at frequencies less than 0.5 Hz is a reasonable approximation when there is no backrest.

For lateral oscillation with a backrest, the equivalent comfort contour produced overall ride values that declined from the expected level of 0.45 m s^{-2} rms as frequency increased from 0.2 to 1.6 Hz. So the discomfort caused by the higher frequencies was underestimated. When the 0.4 Hz high-pass filter was used, as advocated in BS 6841 (BSI 1987), the contour was at a more similar level from 0.2 to 1 Hz. The main contributor to the overall ride value was the lateral oscillation at the seat, with a small contribution from the backrest and little contribution from motion of the feet.

For roll oscillation without a backrest, the equivalent comfort contour produced an overall ride value that was influenced by different components at different frequencies. From 0.2 to 0.5 Hz the seat lateral acceleration was dominant, whereas at frequencies greater than 0.8 Hz the roll acceleration was dominant. The rss of all components is roughly constant, at about the level of the reference (i.e., 0.2 Hz 0.45 m s^{-2} rms lateral oscillation with a rigid seat and backrest), but with a slight trough between 0.63 and 0.8 Hz at the transition between dominance by the seat lateral acceleration and dominance by the seat roll acceleration. The two components at the feet contributed little to the rss of components.

For roll oscillation with a backrest, the seat lateral acceleration again dominates at low frequencies (less than 0.63 Hz) and the seat roll acceleration dominates at high frequencies (greater than 0.63 Hz). The lateral acceleration at the seat back makes a small contribution at low frequencies whereas the tangential acceleration at the seat back makes a small contribution at high frequencies. There is an overall decline in the root-sums of squares of all the components, indicating that the standardised method underestimated the discomfort caused by the higher frequencies relative to the discomfort caused by the 0.2 Hz reference condition.

Within the range of conditions studied here, with and without a backrest, the standardised methods judge the lateral oscillation at all frequencies based on the lateral acceleration on the seat. Ignoring the high-pass filter at 0.4 Hz, at frequencies less than 0.5 Hz, the standardised methods also judge the roll oscillation based on the lateral acceleration (i.e. the acceleration arising from the angle of roll of the seat—g sin θ). At frequencies greater than 0.5 Hz, the roll acceleration is the dominant component, both with and without a backrest.

The standards were evolved from studies with vibration in individual axes, so if one axis dominates they should pro-

vide a good prediction. However, this assumes that the weightings for different axes are given the appropriate relative weight and that relative motion between axes, and the phase of any relative motion, does not influence discomfort. In practice, the relative motion between inputs (e.g. between the seat and the feet) can have a large influence on discomfort (Jang and Griffin, 1999, 2000).

The overall ride values obtained with roll oscillation are influenced by the W_e frequency weighting, but the W_e weighting has not been well developed at low frequencies. The rotational acceleration weighting, W_e, is independent of frequency at frequencies less than 1.0 Hz. To maintain a constant weighted value of roll acceleration, the angle of roll must increase in inverse proportion to the square of the frequency, requiring angles of roll that increase by 12 dB per octave with decreasing frequency. These increasing angles of roll correspond to similarly increasing magnitudes of lateral acceleration in the plane of the seat (i.e. g sin θ) as the frequency decreases. Fig. 2 shows that lateral acceleration corresponding to conditions giving equivalent comfort produced by roll declined at almost 12 dB per octave between 0.2 and 1.6 Hz when using a backrest but at less than 6 dB per octave when not using a backrest. This indicates that the frequency weighting W_e, if used as the sole indicator of discomfort caused by roll, gives insufficient weight to the low frequencies studied here, especially when there is no backrest, as may be seen in Fig. 6. However, the contribution of the low frequencies to discomfort may be predicted from the translational acceleration in the plane of the seat caused by the gravity vector.

V. CONCLUSIONS

At frequencies less than 0.4 Hz, acceleration in the plane of the seat is a useful predictor of discomfort, irrespective of whether the acceleration is caused by lateral oscillation or roll through the gravitational vector. At frequencies greater than 0.4 Hz it is necessary to know whether the acceleration is due to rotation or translation: acceleration in the plane of the seat caused by roll produces greater discomfort than the same acceleration caused by lateral acceleration.

A backrest can increase the discomfort associated with lateral acceleration measured in the plane of the seat. At frequencies greater than about 0.4 Hz, with both lateral oscillation and roll oscillation, a full height backrest with harness increased discomfort, with the discomfort increasing rapidly with increasing frequency of acceleration.

British Standards Institution (BSI). (**1973**). "Guide to the safety aspects of human vibration experiments," BS 7085, London.

British Standards Institution (BSI). (**1987**). "Measurement and evaluation of human exposure to whole-body mechanical vibration and repeated shock," BS 6841, London.

Corbridge, C., and Griffin, M. J. (**1986**). "Vibration and comfort: vertical and lateral motion in the range 0.5 to 5.0 Hz," Ergonomics **29**, 249–272.

Fairley, T. E., and Griffin, M. J. (**1990**). "The apparent mass of the seated human body in the fore and aft and lateral directions," J. Sound Vib. **139**, 299–306.

Forta, N. G. (**2005**). "Effect of whole-body vertical vibration on difference thresholds," Proceedings of the 40th United Kingdom Conference on Human Response to Vibration, held at the UK Health and Safety Executive, Liverpool, 17–19 September 2005.

Griffin, M. J. (**1990**). *Handbook of Human Vibration* (Academic, London).

Griffin, M. J., and Whitham, E. M. (**1978**). "Individual variability and its effect on subjective and biodynamic response to whole-body vibration," J. Sound Vib. **58**, 239–250.

Gunston, T. P. G. (**2004**). "Influence of a back restraint on the lateral apparent mass of the seated human body in response to lateral and roll motions," Proceedings of the 39th United Kingdom Conference on Human Response to Vibration, held at RMS Vibration Test Laboratory, Ludlow Shropshire, 17–19 September 2004.

International Standards Organisation. (**1997**). "Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration," ISO 2631-1, Geneva.

Jang, H. K., and Griffin, M. J. (**1999**). "The effect of phase of differential vertical vibration at the seat and feet on discomfort," J. Sound Vib. **223**, 785–794.

Jang, H. K., and Griffin, M. J. (**2000**). "Effect of phase, frequency, magnitude and posture on discomfort associated with differential vertical vibration at the seat and feet," J. Sound Vib. **229**, 273–286.

Miwa, T. (**1967**). "Evaluation methods for vibration effects. Part 1," Ind. Health **5**, 183–205.

Morioka, M. M., and Griffin, M. J. (**2000**). "Difference thresholds for intensity perception of whole-body vertical vibration: Effect of frequency and magnitude," J. Acoust. Soc. Am. **107**, 620–624.

Paddan, G. S., and Griffin, M. J. (**1988**). "The transmission of translational seat vibration to the head-II. Horizontal seat vibration," J. Biomech. **21**, 199–206.

Paddan, G. S., and Griffin, M. J. (**1994**). "Transmission of roll and pitch seat vibration to the head," Ergonomics **37**, 1513–1531.

Parsons, K. C., and Griffin, M. J. (**1978**). "Effect of the position of the axis of rotation on the discomfort caused by whole-body roll and pitch vibrations of seated persons," J. Sound Vib. **58**, 127–141.

Parsons, K. C., and Griffin, M. J. (**1982**). "Vibration and comfort—2 rotational seat vibration," Ergonomics **25**, 631–644.

Pheasant, S. (**1996**). *Bodyspace - anthropometry, ergonomics and the design of work*, 2nd ed. (Taylor and Francis, London).

Rao, B. K. N., and Jones, B. (**1978**). "Equal sensation study of seated subjects in three translational modes," Ergonomics **21**, 123–134.

Shoenberger, R. W. (**1975**). "Subjective response to very low-frequency vibration," Aviat Space Environ. Med. **46**, 785–790.

Stevens, S. S. (**1975**). *Psychophysics—Introduction to its Perceptual, Neural, and Social Prospects*, transaction Ed. (Transaction Publishers, New Brunswick, NJ).

Frequency- and time-domain asymptotic fields near the critical cone in fluid-fluid configuration

Won-seok Lihh[a)]

Global Communication Technology (GCT) Research, Inc., Seoul, Korea

Near the critical cone of a point source placed at the interface between two half-space fluid media, investigation is made of the asymptotic fields in the frequency domain and their synthetic wave forms in the time domain. While the leading-order (and the head-wave) components give a good description of the true fields well off the critical cone, the uniform asymptotic (UA) analysis has to be made for the approximation near the critical cone. The UA analysis splits into two cases, depending on the medium densities and wave speeds. For Case 1, the UA1 approximation is employed that takes into account the proximity of the stationary-phase point to the branch point. In Case 2, the UA2 approximation is employed in which consideration is also given to the proximity of the stationary point to the pole and to the combined effect of the stationary point, the branch point, and the pole. The validities of the asymptotic fields are checked in the time domain by comparing the asymptotic field wave forms against the wave-number-synthetic wave forms. The UA fields show good accuracy and causal behaviors, with the causality of the UA2 fields previously unreported in the literature. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2714911]

I. INTRODUCTION

Asymptotic analysis of the frequency-domain fields has been a subject of continued study for stratified structures. In the presence of a single planar interface, the (reflected) far-zone or high-frequency fields have received particular attention. The fields can be electromagnetic (Annan, 1973; Engheta *et al.* 1982; Chew, 1990, Sec. 2.6; Felsen and Marcuvitz, 1994, Sec. 5.5) or acoustic/elastic (Stickler, 1976; Brekhovskikh, 1980, Chap. 4; Brekhovskikh and Godin, 1992, Chap. 1; Frisk, 1994, Chap. 4; Gridin, 1998). In this paper, investigation is made of the far-zone acoustic fields for a mass-ejecting point source placed at the perfect (immiscible and slippery) interface between two half-space fluid media. Despite the simplicity of the physical configuration, the asymptotic analysis is of some difficulty around the critical angle (that defines the critical cone). While previous uniform asymptotic (UA) analyses [by, e.g., Borovikov (1994, Sec. 5.5)] apply to the case of ordinary but incomplete space of medium parameters, the UA analysis in this paper covers the full parametric space, extending the scope of the UA approximation. The interface-source problem can be extended to a reflection problem without difficulty. The reflection problem involves a direct field and a reflected field, the latter being the sum of a perfectly reflected field and the remaining field. The direct and the perfectly reflected fields have well-known simple expressions, while the remaining field is directly related to the topic in this paper.

For asymptotic analysis, the fields in the space domain are first expressed in terms of their counterparts in the wave number domain (Sec. II). The stationary-phase point in the complex plane of the lateral wave number (k) determines the leading-order far fields, which can be expressed in simple forms that do not contain mathematical special functions (Sec. III). While the leading-order components give a good description of the true fields at (zenithal) angles well off the critical cone, they fail to approximate the fields near the critical cone. Including the head-wave components (Sec. V), after the interface-factor expansion (Sec. IV), can improve the accuracy well outside the critical cone. But discrepancy is still present near the critical cone, at which the head-wave components become infinite in magnitude. Accuracy cannot be claimed if the stationary and branch points are treated isolatedly.

For an approximation uniformly valid through the critical cone, the UA analysis is made on the combined effect of the stationary and branch points. The far fields are then described by the UA1 formula (Sec. VI A) expressed in terms of the parabolic cylinder functions of orders 1/2 and 3/2 (Borovikov, 1994, Sec. 2.4); or equivalently, in terms of the confluent hypergeometric function. The UA1 approximation, however, is valid only for Case 1 in this paper. For Case 2 (Sec. VI B) which is of rather less occurrence, attention must also be paid to the Sommerfeld pole (Chew, 1990, p. 99; Chew and Kong, 1982) which is unavoidably introduced in the interface-factor expansion. This pole is a removable singularity, since it is originally absent on the proper Riemann sheet. However, the pole comes into effect when its contributions are not canceled out in an asymptotic approximation; it causes an erroneous singular behavior of the UA1 formula at an angle inside the critical cone. To nullify the contributions of the pole, consideration must also be given to the combined effect of the stationary point and the pole and to the combined effect of the stationary point, the branch point, and the pole. The UA2 formula is established for Case 2.

Validities of the four asymptotic fields (leading-order, leading-order plus head-wave, UA1, and UA2) are tested in

---

[a)]Electronic mail: wslee@gctsemi.com
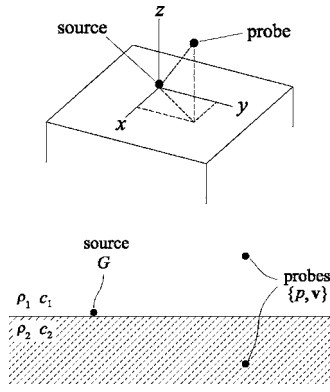
FIG. 1. The half-space radiation problem for a point G source at $(x,y,z) = (0,0,0^+)$. The probes detect the acoustic pressure p and the particle velocity $\mathbf{v}$.

the time domain by comparison against the reference wave forms obtained by the numerical wave-number and frequency synthesis. The wave forms of the UA fields show good accuracy and causal behaviors. The wave-number-frequency synthesis takes more calculation time than the (modified) Cagniard-style evaluation (de Hoop and van der Hijden, 1984, 1985). But, to avoid procedural indirectness, the former is employed to build the reference wave forms. Once the wave-number-synthetic data are calculated in the frequency domain, little time is required for the frequency synthesis into various time windows. Investigation is also made of the dependence of the UA2 fields on the source-to-probe distance along the critical cone.

This paper does not treat the spherical wave components falling off as or faster than the squared reciprocal of the source-to-probe distance. The inhomogeneous surface-wave components (Annan, 1973; Chew, 1990, p. 109) are also not considered, since they are of negligible magnitude well off the interface.

II. TIME-HARMONIC WAVE-NUMBER-SYNTHETIC FIELDS

The physical configuration is shown in Fig. 1. The interface is at $z=0$. The linearized acoustic equations in the physical (space and time) domain are $c^{-2}\partial_t p + \rho \boldsymbol{\nabla} \cdot \mathbf{v} = G$ and $\rho \partial_t \mathbf{v} + \boldsymbol{\nabla} p = 0$ (Kinsler et al., 2000, p. 141), where ρ and c are the medium density and the wave speed, respectively [$\rho = \rho_1$ and $c = c_1$ in the upper medium (medium 1) and $\rho = \rho_2$ and $c = c_2$ in the lower medium (medium 2)]. The fields p and $\mathbf{v}$ are the excess acoustic pressure and the particle velocity, respectively. The source G is the time rate of mass ejection per unit volume. The source generates different amounts of fields depending on whether it is located just above or just below the interface. Hence, let $G = (1\text{ kg})\delta(\mathbf{r})\delta(z-0^\pm)\delta(t)$, where $\mathbf{r} = \mathbf{i}_x x + \mathbf{i}_y y$ and $\delta(\mathbf{r}) = \delta(x)\delta(y)$.

By the composite transform $\mathcal{C} = \mathcal{F}\mathcal{P}$ where $\mathcal{F}\{\ \} = \int_{-\infty}^{\infty}(\ )\exp(-j\omega t)dt$ and $\mathcal{P}\{\ \} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\ )\exp[j(k_x x + k_y y)]dxdy$, one can write the acoustic equations in the domain of the wave numbers (k_x and k_y) and the angular frequency (ω) as

$$\partial_z \dot{p} = -j\omega\rho \dot{\nu}_z, \tag{1}$$

$$\partial_z \dot{\nu}_z = -\frac{\gamma^2}{j\omega\rho}\dot{p} + \frac{1}{\rho}\dot{G}, \tag{2}$$

$$\dot{\nu}_{(x,y)} = \frac{jk_{(x,y)}}{j\omega\rho}\dot{p}, \tag{3}$$

where $\gamma = (k^2 - \omega^2 c^{-2})^{1/2}$ [i.e., $\gamma_n = (k^2 - \omega^2 c_n^{-2})^{1/2}$ in medium n] with $k^2 = k_x^2 + k_y^2$. The fields generated by the subjacent source $\dot{G} = \delta(z-0^-)$ are simply the (ρ_1/ρ_2)-scaled versions of those generated by the superjacent source $\dot{G} = \delta(z-0^+)$, and will not be considered.

Solving Eqs. (1)–(3), one can obtain the acoustic fields in medium n as

$$(\dot{p}_n\ \dot{\nu}_{(x,y)n}\ \dot{\nu}_{zn}) = \frac{1}{\rho_1\rho_n}\frac{e^{-\gamma_n|z|}}{\dfrac{\gamma_1}{\rho_1} + \dfrac{\gamma_2}{\rho_2}}(j\omega\rho_n\ k_{(x,y)}\ \gamma_n\ \mathrm{sgn}(z)). \tag{4}$$

Then the wave-number-synthetic expressions in the space and frequency domain are written as

$$\widetilde{p}_n = \frac{j\omega}{\rho_1}\mathcal{B}_0^{-1}\left\{\frac{e^{-\gamma_n|z|}}{\dfrac{\gamma_1}{\rho_1} + \dfrac{\gamma_2}{\rho_2}}\right\}, \tag{5}$$

$$\widetilde{\nu}_{rn} = \frac{1}{\rho_1\rho_n}\mathcal{B}_1^{-1}\left\{\frac{ke^{-\gamma_n|z|}}{\dfrac{\gamma_1}{\rho_1} + \dfrac{\gamma_2}{\rho_2}}\right\}, \tag{6}$$

$$\widetilde{\nu}_{zn} = \frac{\mathrm{sgn}(z)}{\rho_1\rho_n}\mathcal{B}_0^{-1}\left\{\frac{\gamma_n e^{-\gamma_n|z|}}{\dfrac{\gamma_1}{\rho_1} + \dfrac{\gamma_2}{\rho_2}}\right\}, \tag{7}$$

where $\widetilde{\nu}_{rn} = (x/r)\widetilde{\nu}_{xn} + (y/r)\widetilde{\nu}_{yn}$ and where the wave-number syntheses (the $\mathcal{B}^{-1}$ operations) are defined by

$$\mathcal{B}_{(0,1)}^{-1}\{G(k)\} = \frac{1}{4\pi}\int_{\mathfrak{C}^- \cup \mathfrak{C}^+} G(k)H_{(0,1)}^{(2)}(kr)kdk \tag{8}$$

with $r = |\mathbf{r}| = (x^2 + y^2)^{1/2}$. The functions $H_0^{(2)}(\ )$ and $H_1^{(2)}(\ )$ are the second-kind Hankel functions of orders 0 and 1, respectively. The integration paths $\mathfrak{C}^+$ and $\mathfrak{C}^-$ are, respectively, the first- and third-quadrant parts of the path that runs from left to right extremes through the origin $0+j0$ in the complex k plane. Asymptotic analysis is performed for Eqs. (5)–(7) to obtain the time-harmonic far fields, with the leading-order approximation made in the following section.

III. THE LEADING-ORDER FIELDS

Let $\theta \in (0^\circ, 180^\circ)$ be the zenithal angle satisfying $\cos\theta = z/R$ and $\sin\theta = r/R$ where $R = (r^2 + z^2)^{1/2}$. By replacing the Hankel functions in Eq. (8) with their asymptotic forms (for $r \to \infty$):

$$H_{(0,1)}^{(2)}(kr) \to (1,j)\frac{2e^{j\pi/4}e^{-jkr}}{(2\pi kr)^{1/2}}, \tag{9}$$

one can write

Won-seok Lihh: Fields near critical cone
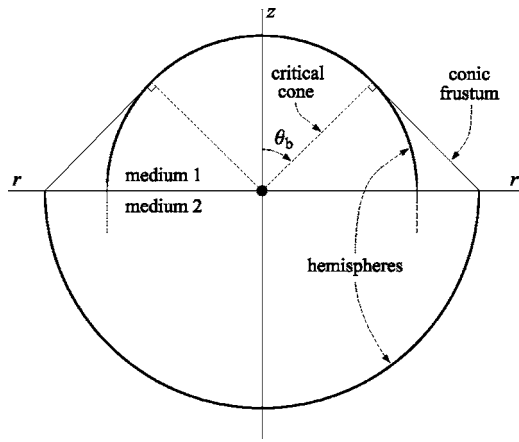
FIG. 2. The wave fronts of the acoustic wave (when $c_1 < c_2$). The head-wave front is on the sidewall of the conic frustum. The critical angle θ_b defines the critical cone.

$$\mathcal{B}_{(0,1)}^{-1}\{F(k)e^{-\gamma_n|z|}\} \sim \frac{(1,j)e^{j\pi/4}}{(2\pi)^{3/2}r^{1/2}}\int_{\mathfrak{C}_k} k^{1/2}F(k)e^{-j\Phi_n(k)R}dk, \quad (10)$$

where $\Phi_n(k) = k\sin\theta + (\omega^2 c_n^{-2} - k^2)^{1/2}|\cos\theta|$. Solving $\Phi_n'(k_{sn}) = 0$ gives the stationary-phase point $k_{sn} = \omega c_n^{-1}\sin\theta$ ($+j0^+$). The integration path $\mathfrak{C}_k$ runs from $k = (k_{sn} - \epsilon) + j0^+$ to $(k_{sn} + \epsilon) + j0^+$, with ϵ being a free positive quantity. Then the following leading-order approximation can be made:

$$\mathcal{B}_{(0,1)}^{-1}\{F(k)e^{-\gamma_n|z|}\} \sim \frac{(j,-1)k_{sn}^{1/2}F(k_{sn})}{2\pi r^{1/2}[-\Phi_n''(k_{sn})R]^{1/2}}e^{-j\Phi_n(k_{sn})R}, \quad (11)$$

where $\Phi_n(k_{sn}) = \omega c_n^{-1}$ and $\Phi_n''(k_{sn}) = -(\omega c_n^{-1}\cos^2\theta)^{-1}$. Equation (A1) has been employed, with

$$I_0^{(0)} = \frac{(2\pi)^{1/2}e^{j\pi/4}}{(2R)^{1/2}}e^{-j\Phi_n(k_{sn})R} \quad (12)$$

from Eq. (A14). Applying Eq. (11) to Eqs. (5)–(7) and expressing the particle velocity in spherical components, the leading-order (LO) fields are obtained as

$$(\tilde{p}_n^{\text{LO}} \quad \tilde{v}_{Rn}^{\text{LO}} \quad \tilde{v}_{\theta n}^{\text{LO}}) = \frac{Q_n}{\rho_1}(1 \quad \eta_n^{-1} \quad 0), \quad (13)$$

where $\eta_n = \rho_n c_n$ and

$$Q_n = \frac{j\omega\rho_n}{2\pi R}\frac{|\cos\theta|}{|\cos\theta| + X_n(\theta)}e^{-j\omega c_n^{-1}R} \quad (14)$$

with $X_n(\theta) = -j(\rho_n/\rho_{\bar{n}})[\sin^2\theta - (c_n^2/c_{\bar{n}}^2)]^{1/2}$. The indices n and $\bar{n}$ are complementarily assigned; that is, $(n,\bar{n}) = (1,2)$ or $(2,1)$. The function $X_n(\theta)$ becomes real and positive when $\sin\theta < c_n/c_{\bar{n}}$. Solving $X_\kappa(\theta_b) = 0$ gives the critical angle

$$\theta_b = 90° + \text{sgn}(c_1 - c_2)\tan^{-1}\frac{(c_{\bar{\kappa}}^2 - c_\kappa^2)^{1/2}}{c_\kappa}, \quad (15)$$

where κ and $\bar{\kappa}$ are such that $c_\kappa = \min\{c_1, c_2\}$ and $c_{\bar{\kappa}} = \max\{c_1, c_2\}$. The critical angle defines the critical cone that forms in medium κ; see Fig. 2.

(a) (b)
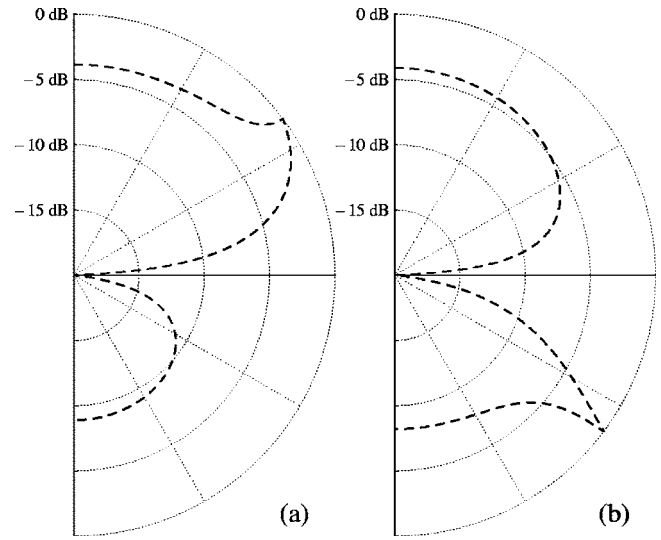
FIG. 3. The radiation patterns of $(1/2)\text{Re}\{(\tilde{v}_R^{\text{LO}})^*\tilde{p}^{\text{LO}}\}$ for (a) the Case 1 parameters and (b) the Case 2 parameters in Table I. The critical angle (θ_b) is approximately 53.1° in (a) and 126.9° in (b).

The leading-order fields have a peak of radiation at $\theta = \theta_b$. See Fig. 3 in which the radiation patterns are shown for the parameters of Cases 1 and 2 in Table I. The peak of Case 1 is near $\theta = 53.1°$ and the peak of Case 2 is near $\theta = 126.9°$. The radiation patterns are calculated by $(1/2)\text{Re}\{(\tilde{v}_{Rn}^{\text{LO}})^*\tilde{p}_n^{\text{LO}}\}$, where the asterisk ($*$) denotes the complex conjugate operation. The pressure $\tilde{p}_n^{\text{LO}}$ and the radial velocity $\tilde{v}_{Rn}^{\text{LO}}$ yield the acoustic impedance η_n.

To check the validity (fidelity) of a leading-order approximation, the transient wave forms are numerically calculated for $\tilde{p}_n^{\text{LO}}$ in Eq. (13) and for the wave-number-synthetic (WS) field in Eq. (5) (say, $\tilde{p}_n^{\text{WS}}$), using

$$\mathcal{F}^{-1}\{\tilde{g}(\omega)(\ )\} = \frac{1}{\pi}\int_0^\infty \text{Re}\{\tilde{g}(\omega)(\ )e^{j\omega t}\}d\omega, \quad (16)$$

where $\tilde{g}(\omega)$ is the spectrum of a smooth source signal $g(t)$. In this paper, it is assumed that $g(t) = 10^{-6}\delta_{(10\ \mu s)}(t)$ where $\delta_T(t) = (\pi^{1/2}T)^{-1}\exp(-t^2/T^2)$, so that $\tilde{g}(\omega) = 10^{-6}\tilde{\delta}_{(10\ \mu s)}(\omega)$ where $\tilde{\delta}_T(\omega) = \exp(-\omega^2 T^2/4)$. Around the critical angles of Cases 1 and 2, respectively, Figs. 4 and 5 compare the wave forms of $\tilde{p}_\kappa^{\text{LO}}$ (dashed lines) with those of $\tilde{p}_\kappa^{\text{WS}}$ (solid lines) at $R = 1$ m; $\kappa = 1$ in Fig. 4 and $\kappa = 2$ in Fig. 5. The wave forms in Fig. 4 are in good agreement, though there are small mismatches at $\theta = \theta_b + 24°$ and $\theta_b + 16°$ caused by the head-wave component and small mismatches near the critical angle (at $\theta = \theta_b + 8°$, $\theta_b + 4°$, $\theta_b + 2°$, θ_b, and $\theta_b - 2°$). The head-wave component is associated with the branch point $k_b = \omega c_{\bar{\kappa}}^{-1}$ ($= \omega c_\kappa^{-1}\sin\theta_b$), and the mismatches near the critical angle are associated with the combined effect of

TABLE I. Medium parameters for Cases 1 and 2.

	ρ_1 (kg m^{-3})	c_1 (m s^{-1})	ρ_2 (kg m^{-3})	c_2 (m s^{-1})
Case 1	700	1200	1000	1500
Case 2	1000	1500	2000	1200
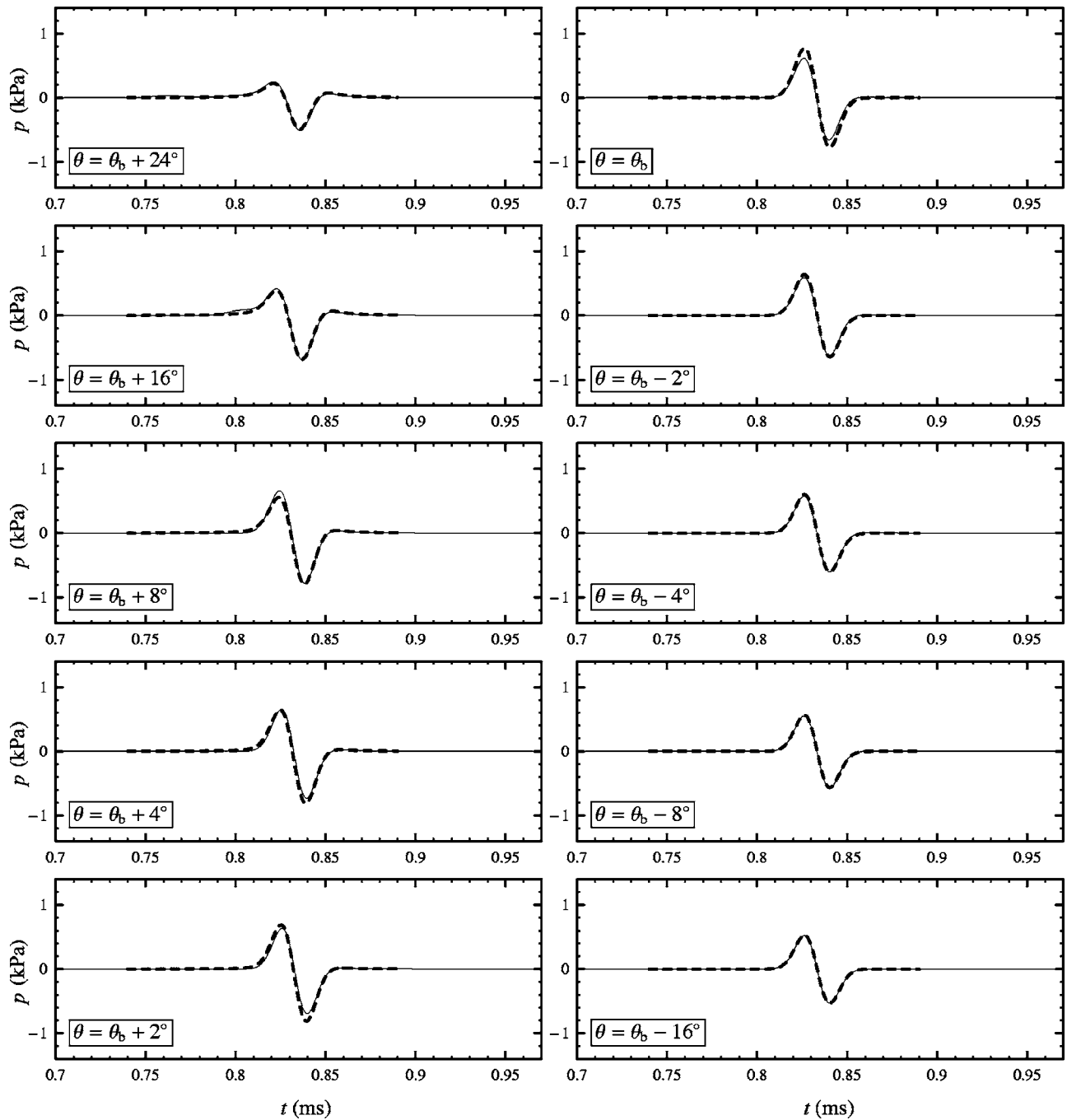
FIG. 4. The wave forms of $\tilde{p}_1^{LO}$ (dashed lines) compared against the wave-number-synthetic wave forms (solid lines) around θ_b ($\approx 53.1°$), for Case 1 ($R = 1$ m).

the stationary and branch points. In Fig. 5 for Case 2, the mismatches are much larger. For better accuracy near the critical angle (in medium κ), the branch point must be taken into account, after expanding the interface factors of the fields in the wave-number domain [say, $(\dot{p}_\kappa \, \dot{v}_{(x,y)\kappa} \, \dot{v}_{z\kappa})/e^{-\gamma_\kappa |z|}$].

IV. INTERFACE-FACTOR EXPANSION

The interface-factor expansion consists of first making the denominator $(\gamma_1/\rho_1)+(\gamma_2/\rho_2)$ in Eq. (4) free of its branch points (similarly as in the process of rationalization) and then employing the relation $(k^2-k_b^2)^{1/2}=(k+k_b)^{1/2}(k-k_b)^{1/2}$ that holds when Re $k>0$ for real k_b. The interface

factor can then be expressed as a sum of two terms each of which has no branch point or contains the branch-point part(s) as multiplicator(s). One term is free of the relevant branch point ($k_b=\omega c_{\bar{\kappa}}^{-1}$) and the other term contains a single factor of this branch point. For example, the interface factor of the field $\dot{p}_\kappa$ from Eq. (4) can be expanded as

$$\frac{\dot{p}_\kappa}{e^{-\gamma_\kappa|z|}} = \frac{\rho_1\rho_2^2}{\rho_\kappa^2-\rho_{\bar{\kappa}}^2}\frac{j\omega}{k^2-\omega^2\Upsilon^{-1}}\left[-\frac{(k^2-\omega^2 c_\kappa^{-2})^{1/2}}{\rho_\kappa}\right.$$
$$\left.+\frac{(k+\omega c_{\bar{\kappa}}^{-1})^{1/2}}{\rho_{\bar{\kappa}}}(k-\omega c_{\bar{\kappa}}^{-1})^{1/2}\right], \qquad (17)$$

where

Won-seok Lihh: Fields near critical cone
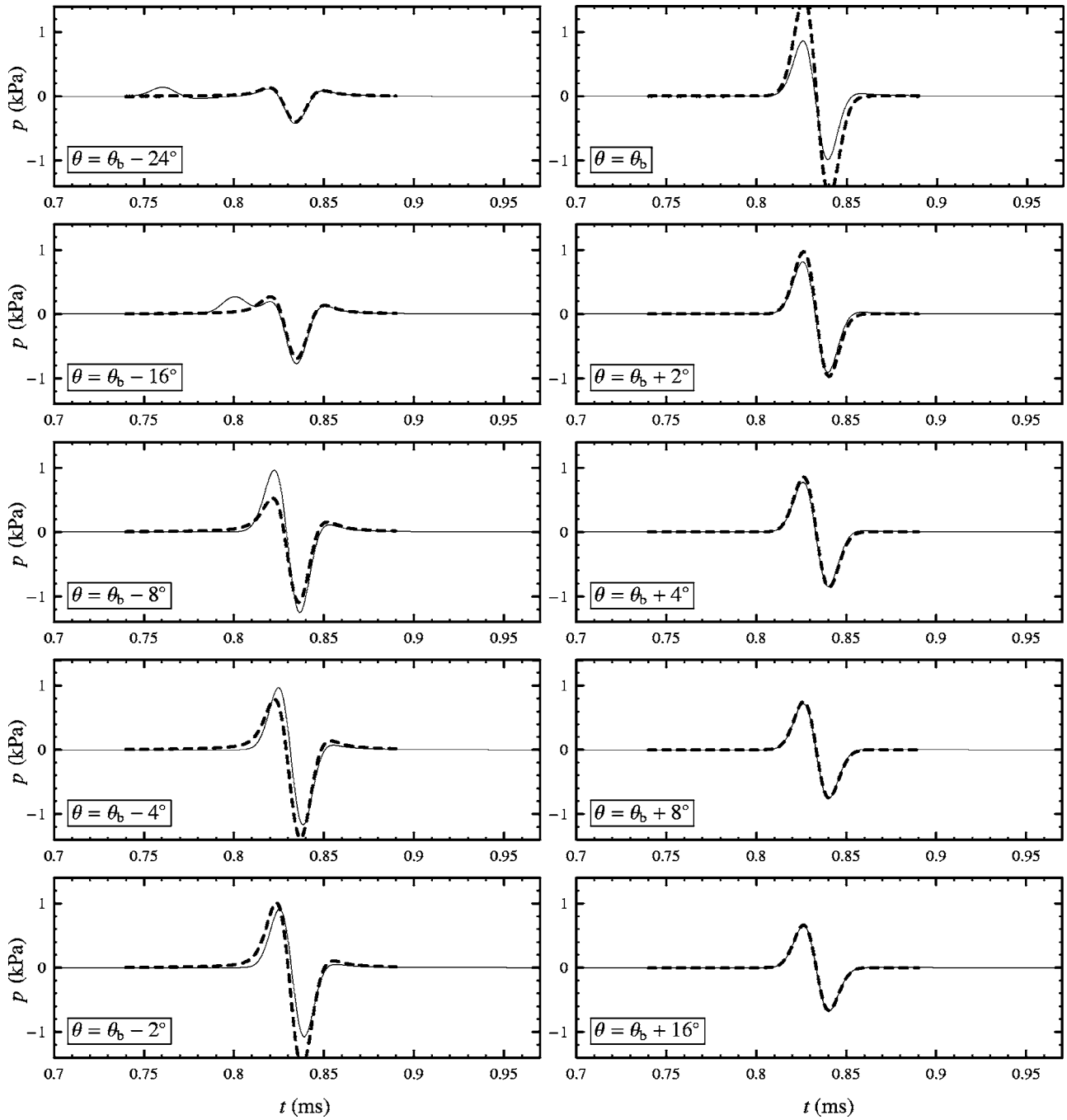
FIG. 5. The wave forms of $\tilde{p}_2^{\text{LO}}$ (dashed lines) compared against the wave-number-synthetic wave forms (solid lines) around θ_b ($\approx 126.9°$), for Case 2 (R =1 m).

$$\Upsilon = \frac{(\rho_2^2 - \rho_1^2)c_1^2 c_2^2}{\rho_2^2 c_2^2 - \rho_1^2 c_1^2}. \tag{18}$$

The velocity components, $\dot{v}_{r\kappa}$ and $\dot{v}_{z\kappa}$, can be similarly expanded. Then, by Eqs. (8) and (9), the far fields in the frequency domain can be calculated via

$$\frac{e^{j\pi/4}}{(2\pi)^{3/2}r^{1/2}}\int_{\mathfrak{C}_k}[\mathsf{M}(k) + j\mathsf{N}(k)(k-k_b)^{1/2}]e^{-j\Phi_\kappa(k)R}dk, \tag{19}$$

where $\mathfrak{C}_k$ runs from $k=(k_{s\kappa}-\epsilon)+j0^+$ to $(k_{s\kappa}+\epsilon)+j0^+$. The functions $\mathsf{M}(k)$ and $\mathsf{N}(k)$ are listed in Table II for each acoustic field.

TABLE II. Expression of $\mathsf{M}(k)$ and $\mathsf{N}(k)$.

	$\mathsf{M}(k)$	$\mathsf{N}(k)$
$\tilde{p}_\kappa$	$\dfrac{\rho_{\bar\kappa}}{\rho_1}\dfrac{\rho_1\rho_2}{\rho_\kappa^2-\rho_{\bar\kappa}^2}\dfrac{\omega k^{1/2}(\omega^2 c_\kappa^{-2}-k^2)^{1/2}}{k^2-\omega^2\Upsilon^{-1}}$	$\dfrac{\rho_\kappa}{\rho_1}\dfrac{\rho_1\rho_2}{\rho_\kappa^2-\rho_{\bar\kappa}^2}\dfrac{\omega k^{1/2}(k+\omega c_{\bar\kappa}^{-1})}{k^2-\omega^2\Upsilon^{-1}}$
$\tilde{v}_{r\kappa}$	$\dfrac{\rho_{\bar\kappa}/\rho_\kappa}{\rho_1}\dfrac{\rho_1\rho_2}{\rho_\kappa^2-\rho_{\bar\kappa}^2}\dfrac{k^{3/2}(\omega^2 c_\kappa^{-2}-k^2)^{1/2}}{k^2-\omega^2\Upsilon^{-1}}$	$\dfrac{1}{\rho_1}\dfrac{\rho_1\rho_2}{\rho_\kappa^2-\rho_{\bar\kappa}^2}\dfrac{k^{3/2}(k+\omega c_{\bar\kappa}^{-1})}{k^2-\omega^2\Upsilon^{-1}}$
$\tilde{v}_{z\kappa}$	$-\dfrac{\rho_{\bar\kappa}/\rho_\kappa}{\rho_1}\dfrac{\rho_1\rho_2}{\rho_2^2-\rho_1^2}\dfrac{k^{1/2}(\omega^2 c_\kappa^{-2}-k^2)}{k^2-\omega^2\Upsilon^{-1}}$	$-\dfrac{1}{\rho_1}\dfrac{\rho_1\rho_2}{\rho_2^2-\rho_1^2}\dfrac{k^{1/2}(\omega^2 c_\kappa^{-2}-k^2)^{1/2}(k+\omega c_{\bar\kappa}^{-1})}{k^2-\omega^2\Upsilon^{-1}}$

Before going on to the UA analysis for the combined effect of the stationary point ($k_{s\kappa}$) and the branch point (k_b), it is worthwhile to consider the isolated contribution of the branch point; that is, the head-wave fields.

V. THE HEAD-WAVE FIELDS

Equation (19) yields the branch-point contribution as

$$\frac{2^{-1/2}\mathsf{N}(k_b)}{2\pi r^{1/2}[\Phi'_\kappa(k_b)R]^{3/2}}e^{-j\Phi_\kappa(k_b)R} \tag{20}$$

(Chew, 1990, Sec. 2.5.2(b); Felsen and Marcuvitz, 1994, Sec. 4.8), which is valid for $\Phi'_\kappa(k_b)>0$, or in other words, valid in the region between the interface and the critical cone. Note the dependence of Eq. (20) on $r^{-1/2}R^{-3/2}$. The head-wave (HW) fields are calculated as

$$(\tilde{p}^{HW}_\kappa \quad \tilde{\nu}^{HW}_{r\kappa} \quad \tilde{\nu}^{HW}_{z\kappa}) = (\eta_\kappa \sec\theta_b \tan\theta_b \quad 1)\frac{\rho_{\bar\kappa}/\rho_\kappa}{\rho_1}$$

$$\times \frac{\exp\left[-j\omega c^{-1}_{\bar\kappa}\left(r+\frac{z}{\tan\theta_b}\right)\right]}{2\pi r^{1/2}(r-z\tan\theta_b)^{3/2}}\tan\theta_b \tag{21}$$

for $|z|<r/|\tan\theta_b|$ in medium κ. The head-wave fields propagate parallel to $\mathbf{i}_\rho\sin\theta_b+\mathbf{i}_z\cos\theta_b$, in which direction the head-wave impedance is η_κ.

Figure 6 compares the wave forms of the leading-order plus head-wave field $\tilde{p}^{LO}_2+\tilde{p}^{HW}_2$ (dashed lines) against the wave-number-synthetic wave forms (solid lines) in Fig. 5. At angles $\theta_b-24°$ and $\theta_b-16°$, the agreement is better than in Fig. 5. As θ approaches θ_b, however, the head-wave component grows to infinity and the agreement becomes worse.

VI. THE UA FIELDS

As seen in Secs. III and V, the individual contributions of the stationary point ($k_{s\kappa}$) and the branch point (k_b) do not give a proper description of the asymptotic fields near the critical angle. Consideration must be given to the combined effect of the stationary and branch points, after writing Eq. (19) in the form

$$\frac{e^{j\pi/4}}{(2\pi)^{3/2}r^{1/2}}\left(\int_{\mathfrak{C}_k}\mathsf{M}(k)+\int_{\mathfrak{C}_k}j\mathsf{N}(k)(k-k_b)^{1/2}\right)e^{-j\Phi_\kappa(k)R}dk. \tag{22}$$

Two UA formulas are derived depending on the value of the parameter Υ [in Eq. (18)]. Let Case 1 correspond to the case $\Upsilon<c^2_\kappa$, and Case 2 to $\Upsilon>c^2_{\bar\kappa}$. The case $c^2_\kappa<\Upsilon<c^2_{\bar\kappa}$ does not occur, since Υ divides the interval $[c^2_\kappa,c^2_{\bar\kappa}]$ externally, in the ratio $\eta^2_\kappa:\eta^2_{\bar\kappa}$ ($=\rho^2_\kappa c^2_\kappa:\rho^2_{\bar\kappa}c^2_{\bar\kappa}$). The relations between the Υ values and the medium parameters are shown by diagrams in Fig. 7. Case 1 corresponds to $(c_1-c_2)(\eta_1-\eta_2)>0$ [i.e., $(c_1-c_2)(\rho_1c_1-\rho_2c_2)>0$], and Case 2 corresponds to $(c_1-c_2)\times(\eta_1-\eta_2)<0$.

A. UA1 approximation for Case 1

In Case 1 ($\Upsilon<c^2_\kappa$), the stationary point $k_{s\kappa}=\omega c^{-1}_\kappa\sin\theta$ [$\in(0,\omega c^{-1}_\kappa)$] does not encounter the pole [of the
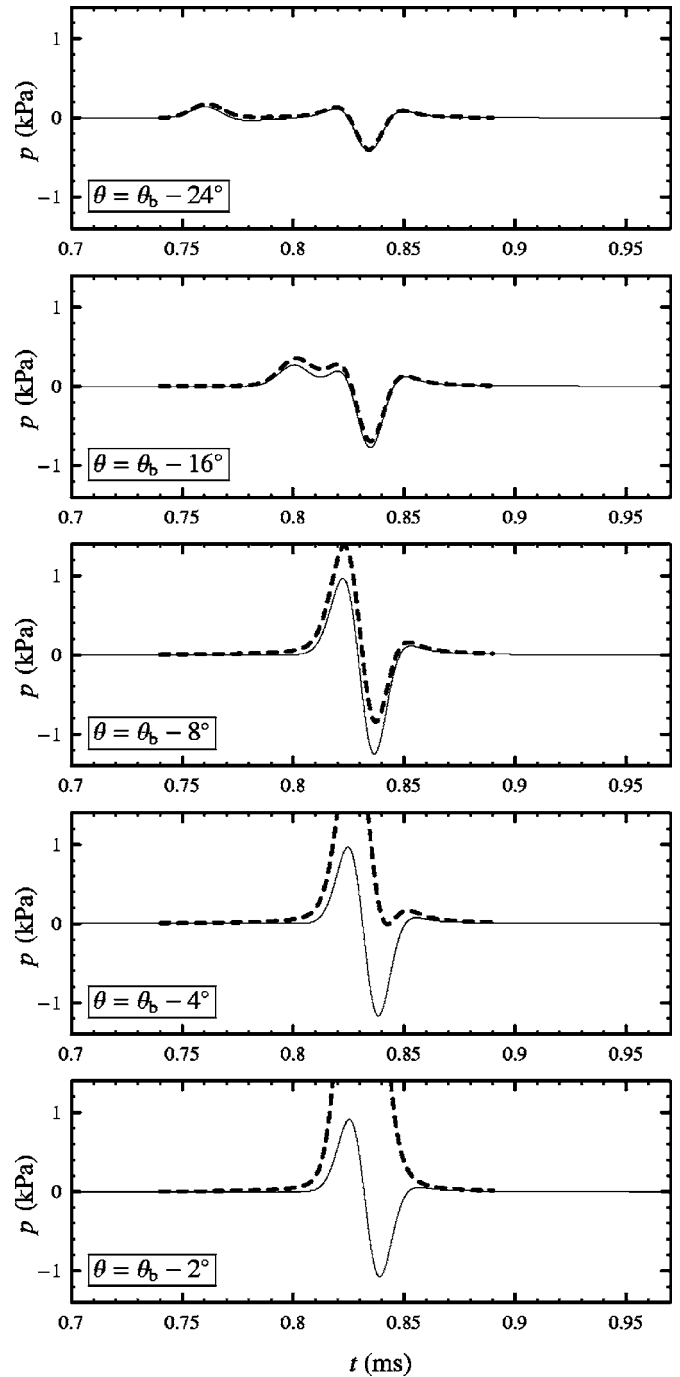
FIG. 6. The wave forms of $\tilde{p}^{LO}_2+\tilde{p}^{HW}_2$ (dashed lines) compared against the wave-number-synthetic wave forms (solid lines) in Fig. 5.

functions $\mathsf{M}(\ )$ and $\mathsf{N}(\ )$] that satisfies $k^2-\omega^2\Upsilon^{-1}=0$ as in Eq. (17). The first integral in Eq. (22) is associated with the stationary point only, while the second integral is associated with the stationary and branch points. The asymptotic fields near $\theta=\theta_b$ can be calculated using the following UA (UA1) formula:

$$\frac{e^{j\pi/4}}{(2\pi)^{3/2}r^{1/2}}\left(\mathsf{M}(k_{s\kappa})\Xi_sI^{(0)}_0+j\mathsf{N}(k_b)\Xi^{3/2}_bI^{(0)}_{1/2}\right.$$

$$\left.+j\frac{\Xi^{3/2}_b\mathsf{N}(k_b)-\Xi^{1/2}_{sb}\Xi_s\mathsf{N}(k_{s\kappa})}{\zeta_b}I^{(0)}_{3/2}\right), \tag{23}$$

which can be obtained from Eqs. (A1) and (A2) by taking
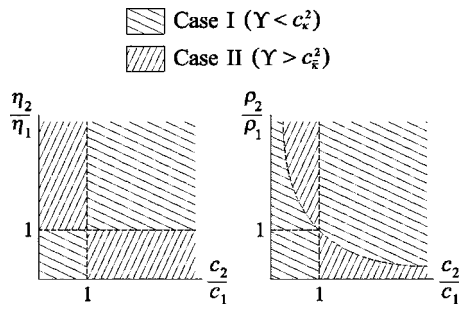
FIG. 7. Diagrams depicting the relations between the Υ values and the medium parameters.

$$(A(k)\ B(k)) = \frac{e^{j\pi/4}}{(2\pi)^{3/2}r^{1/2}}(M(k)\ jN(k)).\qquad(24)$$

The ζ and Ξ quantities in Eq. (23) are given by

$$\zeta_b = \mathrm{sgn}(c_1 - c_2)(2\omega c_\kappa^{-1})^{1/2}\sin\frac{\theta - \theta_b}{2},\qquad(25)$$

$$\Xi_s = (2\omega c_\kappa^{-1})^{1/2}|\cos\theta|,\qquad(26)$$

$$\Xi_b = (2\omega c_\kappa^{-1})^{1/2}\frac{|\cos\theta|}{\cos\dfrac{\theta - \theta_b}{2}},\qquad(27)$$
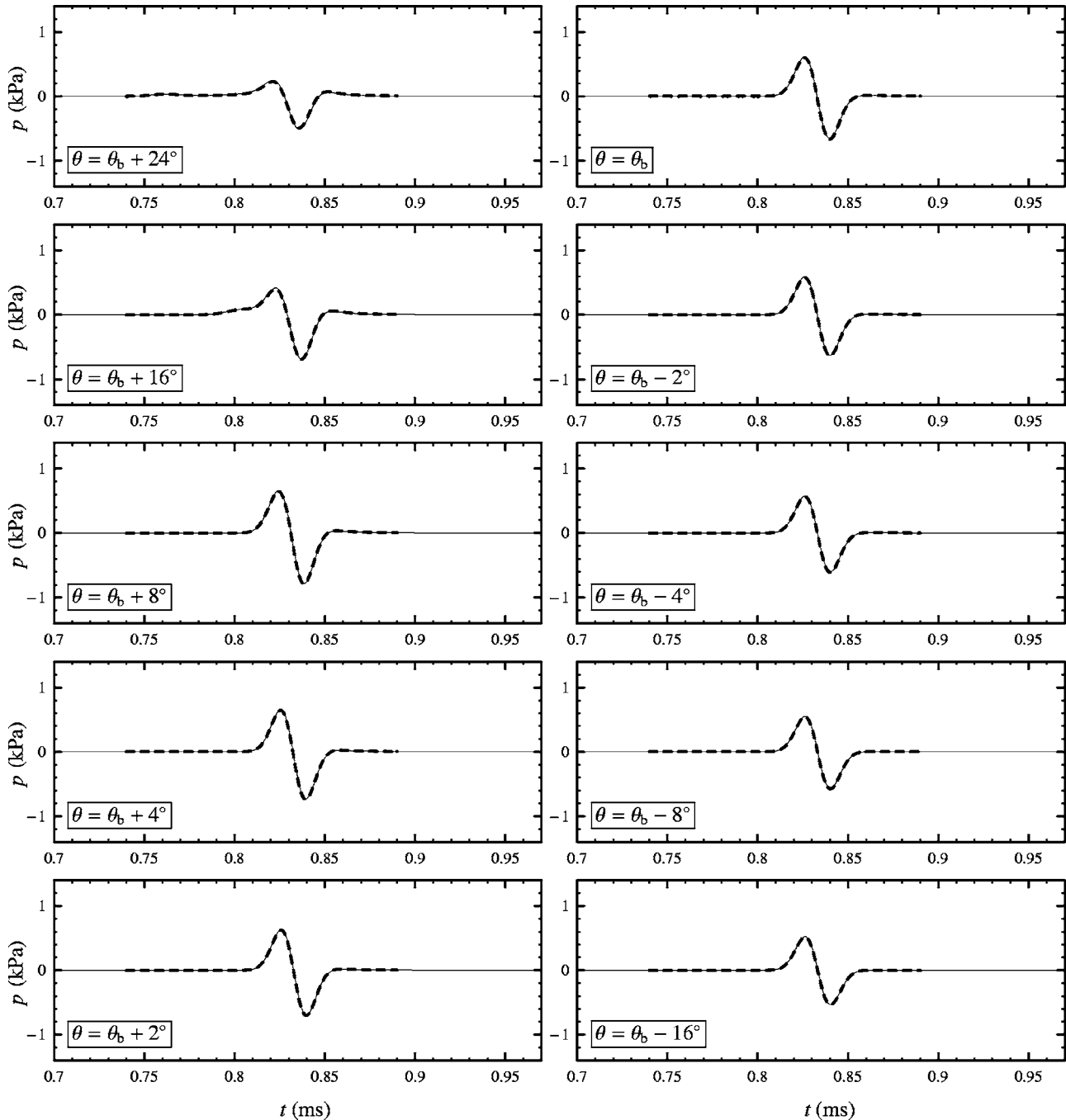
FIG. 8. The wave forms of $\tilde{p}_1^{\mathrm{UA1}}$ (dashed lines) compared against the wave-number-synthetic wave forms (solid lines) in Fig. 4.

$$\Xi_{sb} = (2\omega c_\kappa^{-1})^{1/2} \left| \cos \frac{\theta + \theta_b}{2} \right| \tag{28}$$

from Eqs. (A5) and (A10)–(A12), while the I factors are evaluated from Eq. (A14) as

$$I_0^{(0)} = \frac{(2\pi)^{1/2} e^{j\pi/4}}{(2R)^{1/2}} e^{-j\Phi_\kappa(k_{s\kappa})R}, \tag{29}$$

$$I_{1/2}^{(0)} = \frac{(2\pi)^{1/2} e^{j5\pi/8}}{(2R)^{3/4}} D_{1/2}(w_b) e^{-j(1/2)[\Phi_\kappa(k_{s\kappa})+\Phi_\kappa(k_b)]R}, \tag{30}$$

$$I_{3/2}^{(0)} = \frac{(2\pi)^{1/2} e^{-j5\pi/8}}{(2R)^{5/4}} D_{3/2}(w_b) e^{-j(1/2)[\Phi_\kappa(k_{s\kappa})+\Phi_\kappa(k_b)]R}, \tag{31}$$

where $w_b = e^{j\pi/4}(2R)^{1/2}\zeta_b$ and $\Phi_\kappa(k_b) = \omega c_\kappa^{-1} \cos(\theta - \theta_b)$. The parabolic cylinder functions are calculated from

$$D_\nu(z) = \frac{2^{\nu/2}}{e^{z^2/4}} \left[\frac{\pi^{1/2}}{\Gamma\left(\frac{1-\nu}{2}\right)} \, {}_1F_1\left(-\frac{\nu}{2}, \frac{1}{2}; \frac{z^2}{2}\right) \right.$$
$$\left. - \frac{(2\pi)^{1/2}z}{\Gamma\left(-\frac{\nu}{2}\right)} \, {}_1F_1\left(\frac{1-\nu}{2}, \frac{3}{2}; \frac{z^2}{2}\right) \right] \tag{32}$$

(Gradshteyn and Ryzhik, 1994, p. 1092), where the special functions Γ and ${}_1F_1$ are the gamma function and the confluent hypergeometric function, respectively, of which the numerical values can be obtained using software like MATH-EMATICA (Wolfram, 1999). In Fig. 8 are shown the wave forms of the UA1 field $\tilde{p}_1^{UA1}$ (dashed lines) for the Case 1 parameters in Table I, compared against the wave-number-synthetic wave forms in Fig. 4 (solid lines). Better agreement can be seen than in Fig. 4. Exactly at the critical angle, the stationary point $k_{s\kappa}$ coalesces with the branch point k_b. In this case, the UA1 formula enjoys the asymptoticity of R^{-1}, $R^{-5/4}$, and $R^{-7/4}$.

If $k_{s\kappa}$ is not close to k_b, then Eq. (23) becomes

$$j\frac{M(k_{s\kappa}) + jN(k_{s\kappa})(k_{s\kappa} - k_b)^{1/2}}{2\pi r^{1/2}[-\Phi_\kappa''(k_{s\kappa})R]^{1/2}} e^{-j\Phi_\kappa(k_{s\kappa})R}$$
$$+ \frac{2^{-1/2}N(k_b)e^{-j\Phi_\kappa(k_b)R}}{2\pi r^{1/2}[\Phi_\kappa'(k_b)R]^{3/2}} u(k_{s\kappa} - k_b), \tag{33}$$

according to the asymptotic behaviors of the parabolic cylinder function (for large $|z|$):

$$D_\nu(z) \sim \frac{z^\nu}{e^{z^2/4}} \quad \left(|\arg z| < \frac{3\pi}{4}\right), \tag{34a}$$

$$\sim \frac{z^\nu}{e^{z^2/4}} - \frac{(2\pi)^{1/2}}{\Gamma(-\nu)e^{j\pi\nu}} \frac{e^{z^2/4}}{z^{\nu+1}} \quad \left(-\frac{5\pi}{4} < \arg z < -\frac{\pi}{4}\right) \tag{34b}$$

(Gradshteyn and Ryzhik, 1994, p. 1094; Jeffrey and Zwillinger, 2005) for Eqs. (30) and (31). Note that $\Phi_\kappa'(k_b) > 0$ for $k_{s\kappa} > k_b$ due to the assumption $\Phi_\kappa''(k_{s\kappa}) < 0$. Equation (33) covers Eqs. (11) and (20), giving the components of R^{-1} and R^{-2}.

B. UA2 approximation for Case 2

In Case 2 ($\Upsilon > c_{\bar{\kappa}}^2$), the Sommerfeld pole $k_p = \omega\Upsilon^{-1/2}$ is located in the range $(0, \omega c_\kappa^{-1})$ of the stationary point $k_{s\kappa} = \omega c_\kappa^{-1} \sin\theta$. In this case, the first integral in Eq. (22) is associated with the stationary point and the pole, while the second integral is associated with the stationary point, the branch point, and the pole. The UA1 formula [Eq. (23)] cannot be used for Case 2. See the discrepancy in Fig. 9 between the wave forms of $\tilde{p}_2^{UA1}$ (dashed lines) and the wave-number-synthetic wave forms in Fig. 5 (solid lines). When directly evaluating the integral in Eq. (19), the pole singularities of $M(k)$ and $N(k)$ cancel out each other, since the original wave-number-domain expressions in Eq. (4) are pole-free on the proper Riemann sheet. However, the cancellation is imperfect in Eq. (23), as can be noticed from the expressions of $I_0^{(0)}$, $I_{1/2}^{(0)}$, and $I_{3/2}^{(0)}$ in Eqs. (29)–(31). The dependence of $|I_0^{(0)}|$ on R is different from those of $|I_{1/2}^{(0)}|$ and $|I_{3/2}^{(0)}|$, unless the argument w_b of the parabolic cylinder functions has a sufficiently large magnitude. The ζ and L quantities in Eq. (23) being independent of R, there does not occur an exact cancellation of the pole contributions. Additional measures must be taken to cancel them out.

The right way of analysis for Case 2 begins by rewriting Eq. (22) in the form

$$\frac{e^{j\pi/4}}{(2\pi)^{3/2}r^{1/2}} \left(\int_{\mathfrak{C}_k} M_0(k) + \int_{\mathfrak{C}_k} jN_0(k)(k - k_b)^{1/2} \right.$$
$$\left. + \check{M} \int_{\mathfrak{C}_k} \frac{1}{k - k_p} + j\check{N} \int_{\mathfrak{C}_k} \frac{(k - k_b)^{1/2}}{k - k_p} \right) e^{-j\Phi_\kappa(k)R} dk, \tag{35}$$

where $\check{M} = \text{Res}_{k=k_p}\{M(k)\}$, $\check{N} = \text{Res}_{k=k_p}\{N(k)\}$, and

$$\begin{pmatrix} M_0(k) \\ N_0(k) \end{pmatrix} = \begin{pmatrix} M(k) \\ N(k) \end{pmatrix} - \frac{1}{k - k_p} \begin{pmatrix} \check{M} \\ \check{N} \end{pmatrix}. \tag{36}$$

Note that $\check{N} = \check{M}/(k_b - k_p)^{1/2}$, since the expression $M(k) + jN(k)(k - k_b)^{1/2}$ [in Eq. (19)] is nonsingular at $k = k_p$. The first integral in Eq. (35) involves the stationary point only. The second integral involves the stationary point and the branch point. The third involves the stationary point and the pole. The fourth involves all three points. Then, in approximating Eq. (35), the following UA (UA2) formula can be established from Eqs. (A1) and (A2)–(A4):

$$\frac{e^{j\pi/4}}{(2\pi)^{3/2}r^{1/2}} \left\{ \left(\Xi_s M(k_{s\kappa}) + \frac{\check{M}}{\zeta_p} \right) I_0^{(0)} \right.$$
$$+ j\left(\Xi_b^{3/2} N(k_b) - \frac{\Xi_{bp}^{1/2}\check{N}}{\zeta_b - \zeta_p} \right) I_{1/2}^{(0)}$$
$$+ j\left(\frac{\Xi_b^{3/2} N(k_b) - \Xi_{sb}^{1/2}\Xi_s N(k_{s\kappa})}{\zeta_b} - \frac{\Xi_{bp}^{1/2}\check{N}}{\zeta_p(\zeta_b - \zeta_p)} \right) I_{3/2}^{(0)}$$
$$\left. + \check{M}I_0^{(1)} + j\check{N}\Xi_{bp}^{1/2}I_{1/2}^{(1)} \right\}, \tag{37}$$

by taking

$$(\mathsf{A}(k) \quad \mathsf{B}(k) \quad \mathsf{C}(k) \quad \mathsf{D}(k))$$

$$= \frac{e^{j\pi/4}}{(2\pi)^{3/2}r^{1/2}}(\mathsf{M}_0(k) \quad j\mathsf{N}_0(k) \quad \check{\mathsf{M}} \quad j\check{\mathsf{N}}). \tag{38}$$

The various quantities in Eq. (37) are given by Eqs. (25)–(31) and by

$$\zeta_p = \mathrm{sgn}(c_1 - c_2)(2\omega c_\kappa^{-1})^{1/2}\sin\frac{\theta - \theta_p}{2}, \tag{39}$$

$$\Xi_{sp} = (2\omega c_\kappa^{-1})^{1/2}\left|\cos\frac{\theta + \theta_p}{2}\right|, \tag{40}$$

$$\Xi_{bp} = (2\omega c_\kappa^{-1})^{1/2}\frac{\left|\cos\dfrac{\theta_b + \theta_p}{2}\right|\sin\dfrac{\theta_b - \theta_p}{2}}{\sin\dfrac{\theta - \theta_p}{2} - \sin\dfrac{\theta - \theta_b}{2}}, \tag{41}$$

$$I_0^{(1)} = j2\pi\left[\frac{D_{-1}(w_p)}{(2\pi)^{1/2}}e^{-j(1/2)[\Phi_\kappa(k_{s\kappa}) + \Phi_\kappa(k_p)]R} - e^{-j\Phi_\kappa(k_p)R}\right], \tag{42}$$

$$I_{1/2}^{(1)} = 2\pi u^{1/2}(\zeta_b - \zeta_p)e^{-j\Phi_\kappa(k_p)R}$$
$$+ \frac{e^{j\pi/8}e^{-j\Phi_\kappa(k_{s\kappa})R}}{(2R)^{1/4}}\int_{-\infty}^{\infty}\frac{\xi^{1/2}e^{-(1/2)(\xi - jw_b)^2}d\xi}{\xi - j(w_b + w_p)}, \tag{43}$$

where

$$\theta_p = 90° + \mathrm{sgn}(c_1 - c_2)\tan^{-1}\left(\frac{\rho_\kappa^2(c_{\bar{\kappa}}^2 - c_\kappa^2)}{\rho_\kappa^2 c_\kappa^2 - \rho_{\bar{\kappa}}^2 c_{\bar{\kappa}}^2}\right)^{1/2} \tag{44}$$

[from $k_p(=\omega\Upsilon^{-1/2}) = \omega c_\kappa^{-1}\sin\theta_p$], $\Phi_\kappa(k_p) = \omega c_\kappa^{-1}\cos(\theta - \theta_p)$, $w_p = e^{j\pi/4}(2R)^{1/2}(-\zeta_p)$ and $u^{1/2}(\zeta) = \zeta^{1/2}u(\zeta)$ with $u(\ )$ being the unit-step function. Only the case $\zeta_b > \zeta_p$ is relevant in this paper. The UA2 formula differs from the UA1 formula [in Eq. (23)] by the newly introduced terms that contain $\check{\mathsf{M}}$ and $\check{\mathsf{N}}$.

For numerical evaluation of the integral in Eq. (43), it is more efficient to use a faster-converging expression

$$\int_{-\infty}^{\infty}\frac{(\xi + jw_b)^{1/2}}{\xi - jw_p}e^{-(1/2)\xi^2}d\xi$$
$$+ 2u(-\zeta_b)\int_{-\infty}^{0}\frac{\xi^{1/2}e^{-(1/2)(\xi - jw_b)^2}d\xi}{\xi - j(w_b + w_p)}$$
$$+ j2\pi u(\zeta_p)(jw_b + jw_p)^{1/2}e^{(1/2)w_p^2} \tag{45}$$

obtained by deforming the integration contour in the complex ξ plane. See Fig. 10, in which the wave forms of the UA2 field $\tilde{p}_2^{\mathrm{UA2}}$ (dashed lines) are in good agreement with the wave-number-synthetic wave forms (solid lines), for the Case 2 parameters in Table I. This is in contrast to the poor agreements in Figs. 5, 6, and 9.

While the integral in Eq. (43) may be treated as a "canonical" one (Ciarkowski, 1989), some rare special function must be employed to express this integral generally (Bleistein, 1967; Chew and Kong, 1982). As a special case,

however, one can assume the probe to be placed exactly on the critical cone ($w_b = 0$) and evaluate the integral in terms of well-known special functions. For this evaluation, the integral is first written as

$$\left(\int_0^\infty \xi^{1/2} + \int_{-\infty}^0 j(-\xi)^{1/2}\right)\frac{e^{-(1/2)\xi^2}d\xi}{\xi - j\hat{w}_p}$$
$$= \frac{1 - j}{2}\int_0^\infty \frac{x^{1/4} - \hat{w}_p x^{-1/4}}{x + \hat{w}_p^2}e^{-(1/2)x}dx, \tag{46}$$

where $x = \xi^2$ and $\hat{w}_p = e^{j\pi/4}(2R)^{1/2}[\Phi_\kappa(k_b) - \Phi_\kappa(k_p)]^{1/2}$. For each term of the numerator, the x integral is evaluated using

$$\int_0^\infty \frac{x^{\nu - 1}e^{-\mu x}}{x + \beta}dx = \beta^{\nu - 1}e^{\beta\mu}\Gamma(\nu)\Gamma(1 - \nu, \beta\mu) \tag{47}$$

for $|\arg\beta| < \pi$, $\mathrm{Re}\,\mu > 0$, and $\mathrm{Re}\,\nu > 0$ (Gradshteyn and Ryzhik, 1994, p. 366), where the second Γ function is the incomplete gamma function. Then Eq. (46) is expressed as

$$\frac{\hat{w}_p^{1/2}}{2^{5/2}e^{j\pi/4}}e^{(1/2)\hat{w}_p^2}\left[\Gamma\left(\frac{1}{4}\right)\Gamma\left(-\frac{1}{4}, \frac{1}{2}\hat{w}_p^2\right)\right.$$
$$\left. + \Gamma\left(-\frac{1}{4}\right)\Gamma\left(\frac{1}{4}, \frac{1}{2}\hat{w}_p^2\right)\right]. \tag{48}$$

A physical insight regarding the dependence of the UA2 field on the distance R can be gained by looking at the asymptotic behaviors of the special functions (for $R \to \infty$):

$$D_{-1}(\hat{w}_p) \propto R^{-1/2}, \tag{49}$$

$$\Gamma\left(\frac{1}{4}, \frac{1}{2}\hat{w}_p^2\right) \propto R^{-3/4}, \tag{50}$$

$$\Gamma\left(-\frac{1}{4}, \frac{1}{2}\hat{w}_p^2\right) \propto R^{-5/4} \tag{51}$$

according to Eq. (34) and $\Gamma(a, z) \sim z^{a-1}e^{-z}$ (Abramowitz and Stegun, 1972, p. 263). It turns out that the UA2 formula [Eq. (37)] enjoys the asymptoticity of R^{-1}, $R^{-5/4}$, and $R^{-7/4}$ similar to the UA1 formula [Eq. (23)], while the terms of $R^{-1/2}$ resulting from $I_0^{(1)}$ and $I_{1/2}^{(1)}$ cancel out. Although it can be shown that the UA2 formula tends to the UA1 formula at extreme distances on the critical cone, the proof is rather lengthy and is not presented due to space limitations.

VII. CONCLUSION

Investigation has been made of the frequency-domain asymptotic fields and their time-domain wave forms, for a mass-ejecting source placed at the interface between half-space fluid media. The UA approximation has been made near the critical cone, where the leading-order and head-wave fields fail to give a proper description of the true fields. Depending on the densities and wave speeds of the upper and lower media, the UA problem is split into two cases (Cases 1 and 2). For Case 1, it is sufficient to take into account the combined effect of the stationary and branch points, and the wave fields are well described by the UA1 formula. In Case 2, consideration has also been given to the
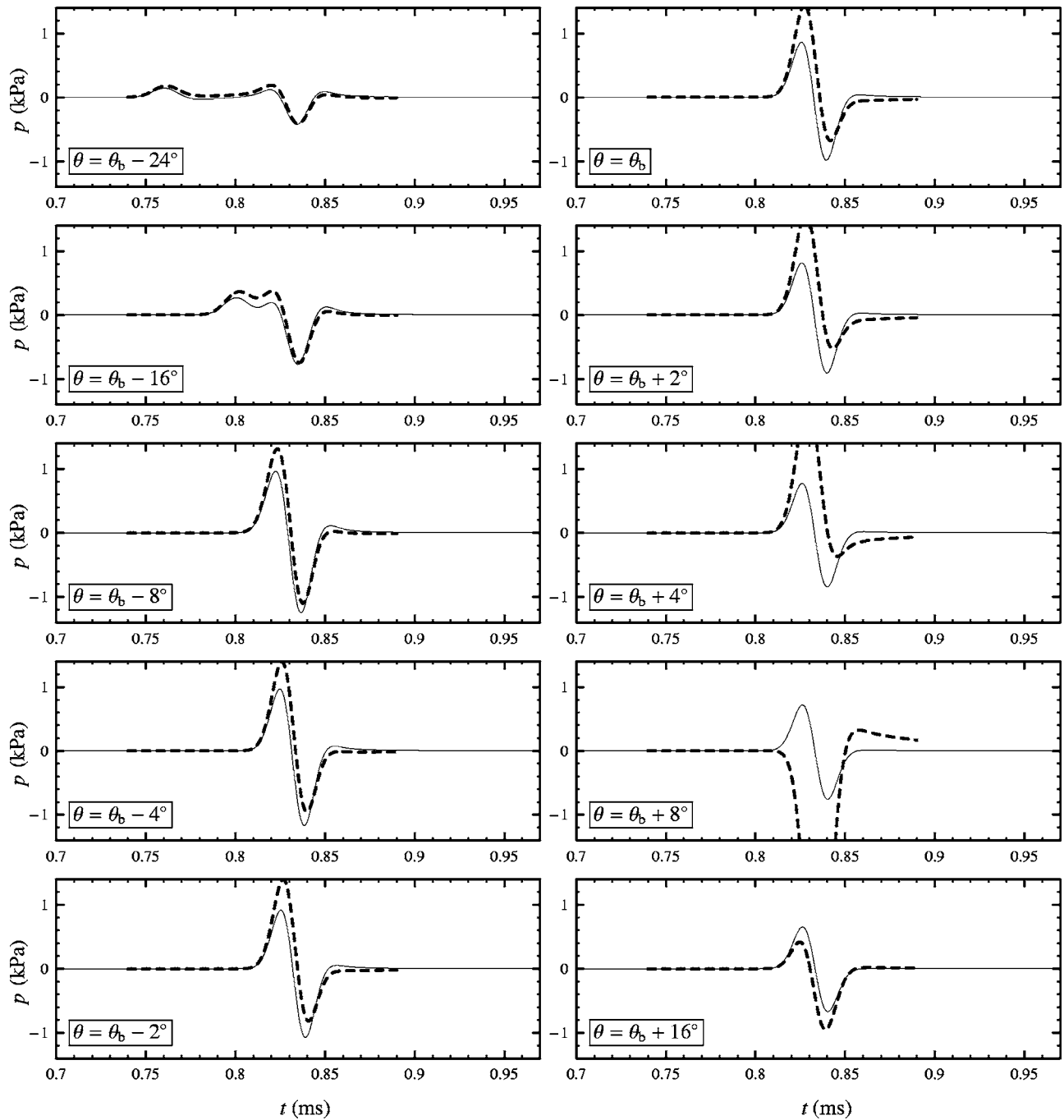
FIG. 9. The wave forms of $\tilde{p}_2^{\mathrm{UA1}}$ (dashed lines) compared against the wave-number-synthetic wave forms (solid lines) in Fig. 5.

combined effect of the stationary point and the pole and to the combined effect of the stationary point, the branch point, and the pole. The Case 2 fields are described by the UA2 formula. The validities of the UA formulas have been verified in the time domain by comparison against the wave-number-synthetic wave forms. Although the general causality of the UA fields has not been theorematically proved in this paper, the frequency-synthesized transients of the UA fields show good causal behaviors.

It might be argued that the special functions and the integral in the UA formulation do not afford immediate physical insights like those of the leading-order approximation. Depending on the spatial region of interest, however,

there exist special cases in which the integral can be evaluated and the special functions involved can be replaced by their asymptotic forms or constant quantities, giving simpler expressions of the UA fields. On the critical cone, especially, the dependence of the UA2 fields on the source-to-probe distance has been revealed. The UA formulation in this paper can serve as a basis for obtaining more definite formulas on the critical cone, if appropriate limits are evaluated for all the present ζ and L quantities.

APPENDIX: ASYMPTOTIC APPROXIMATIONS

For the exponential factor $e^{-j\Phi(k)R}$, let $k_s = k_s(\theta)$ be the positive real stationary-phase point satisfying $\Phi'(k_s) = 0$ and

2664   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Won-seok Lihh: Fields near critical cone
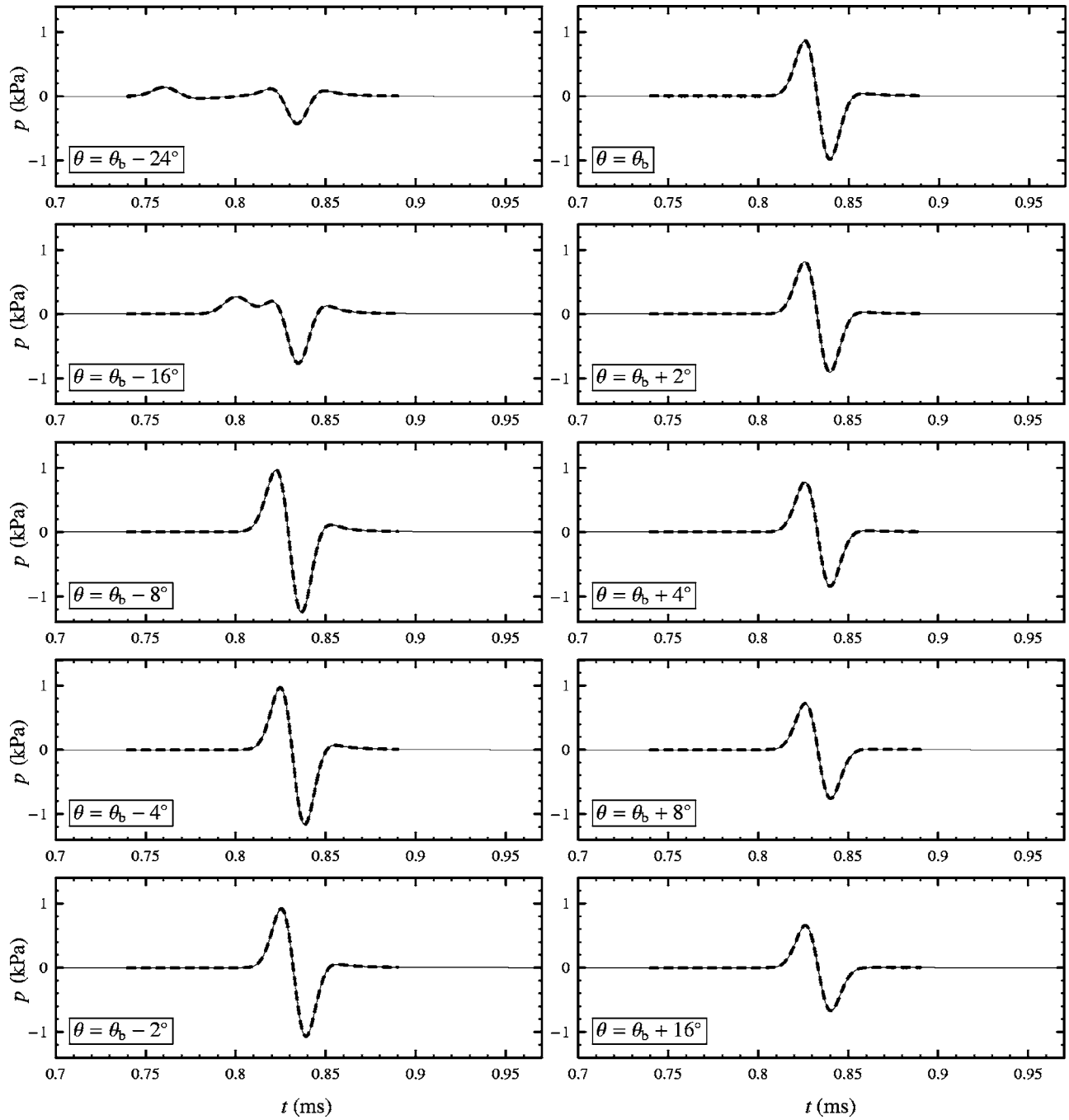
FIG. 10. The wave forms of $\tilde{p}_2^{\mathrm{UA2}}$ (dashed lines) compared against the wave-number-synthetic wave forms (solid lines) in Fig. 5.

$\Phi''(k_s) < 0$. Various asymptotic analyses (Bleistein and Handelsman, 1986, Sec. 9.4; Ciarkowski, 1989; Brekhovskikh and Godin, 1992, Secs. A.2 and A.3; Borovikov, 1994, Secs. 1.2, 2.2, and 2.4; Gallop and Hron, 1999; Wong, 2001, Secs. II.3, VII.2, and VII.3) are reformulated into the following four formulas:

$$\int_{\mathfrak{C}_k} \mathsf{A}(k)e^{-j\Phi(k)R}dk \sim \mathsf{A}_s L_s I_0^{(0)}, \tag{A1}$$

$$\int_{\mathfrak{C}_k} \mathsf{B}(k)(k-k_b)^{1/2}e^{-j\Phi(k)R}dk \sim \mathsf{B}_b L_b^b I_{1/2}^{(0)}$$
$$+ \frac{\mathsf{B}_b L_b^b - \mathsf{B}_s L_s^b}{\zeta_b}I_{3/2}^{(0)}, \tag{A2}$$

$$\int_{\mathfrak{C}_k} \mathsf{C}(k)\frac{e^{-j\Phi(k)R}}{k-k_p}dk \sim \mathsf{C}_p L_p^p I_0^{(1)} + \frac{\mathsf{C}_p L_p^p - \mathsf{C}_s L_s^p}{\zeta_p}I_0^{(0)}, \tag{A3}$$

$$\int_{\mathfrak{C}_k} \mathsf{D}(k)\frac{(k-k_b)^{1/2}}{k-k_p}e^{-j\Phi(k)R}dk \sim \mathsf{D}_p L_p^{bp} I_{1/2}^{(1)}$$
$$+ \frac{\mathsf{D}_b L_b^{bp} - \mathsf{D}_p L_p^{bp}}{\zeta_b - \zeta_p}I_{1/2}^{(0)} + \left(\frac{\mathsf{D}_b L_b^{bp} - \mathsf{D}_s L_s^{bp}}{\zeta_b(\zeta_b - \zeta_p)} \right.$$
$$\left. - \frac{\mathsf{D}_p L_p^{bp} - \mathsf{D}_s L_s^{bp}}{\zeta_p(\zeta_b - \zeta_p)} \right) I_{3/2}^{(0)}, \tag{A4}$$

where $\mathsf{A}_s = \mathsf{A}(k_s)$, $\mathsf{B}_{(b,s)} = \mathsf{B}(k_{(b,s)})$, $\mathsf{C}_{(p,s)} = \mathsf{C}(k_{(p,s)})$, and $\mathsf{D}_{(b,p,s)} = \mathsf{D}(k_{(b,p,s)})$. The integration path $\mathfrak{C}_k$ runs from $k=(k_s$

$-\epsilon)+j0^+$ to $(k_s+\epsilon)+j0^+$, with ϵ being a free positive quantity. The ζ and L quantities, which are independent of R, are given by

$$\zeta_{(b,p)} = \left[\Phi(k_s) - \Phi(k_{(b,p)})\right]^{1/2} \mathrm{sgn}(k_{(b,p)} - k_s), \tag{A5}$$

$$L_s = \Xi_s, \tag{A6}$$

$$(L_b^b \; L_s^b) = (\Xi_b^{3/2} \; \Xi_{sb}^{1/2}\Xi_s), \tag{A7}$$

$$(L_p^p \; L_s^p) = \left(1 \; \frac{\Xi_s}{\Xi_{sp}}\right), \tag{A8}$$

$$(L_p^{bp} \; L_b^{bp} \; L_s^{bp}) = \left(\Xi_{bp}^{1/2} \; \frac{\Xi_b^{3/2}}{\Xi_{bp}} \; \frac{\Xi_{sb}^{1/2}\Xi_s}{\Xi_{sp}}\right), \tag{A9}$$

where

$$\Xi_s = \left(\frac{-2}{\Phi''(k_s)}\right)^{1/2}, \tag{A10}$$

$$\Xi_b = \frac{-2\zeta_b}{\Phi'(k_b)} = \frac{2\left[\Phi(k_s) - \Phi(k_b)\right]^{1/2}}{|\Phi'(k_b)|}, \tag{A11}$$

$$\Xi_{s(b,p)} = \frac{k_s - k_{(b,p)}}{-\zeta_{(b,p)}} = \frac{|k_s - k_{(b,p)}|}{\left[\Phi(k_s) - \Phi(k_{(b,p)})\right]^{1/2}}, \tag{A12}$$

$$\Xi_{bp} = \frac{k_b - k_p}{\zeta_b - \zeta_p}. \tag{A13}$$

The I factors in Eqs. (A1)–(A4) are given by

$$I_a^{(n)} = e^{-j\Phi(k_s)R} \int_{C_\zeta} \frac{(\zeta - \zeta_b)^a}{(\zeta - \zeta_p)^n} e^{j\zeta^2 R} d\zeta, \tag{A14}$$

where the path C_ζ runs left to right extremes, parallel to the real ζ axis with a small positive imaginary part.

Abramowitz, M., and Stegun, I. A. (**1972**). *Handbook of Mathematical Functions* (Dover, Mineola, NY).

Annan, A. P. (**1973**). "Radio interferometry depth sounding. 1. Theoretical discussion," Geophysics **38**, 557–580.

Bleistein, N. (**1967**). "Uniform asymptotic expansions of integrals with many nearby stationary points and algebraic singularities," J. Math. Mech. **17**, 533–559.

Bleistein, N., and Handelsman, R. A. (**1986**). *Asymptotic Expansions of Integrals* (Dover, New York).

Borovikov, V. A. (**1994**). *Uniform Stationary Phase Method* (IEE Press, London).

Brekhovskikh, L. M. (**1980**). *Waves in Layered Media*, 2nd ed. (Academic, New York).

Brekhovskikh, L. M., and Godin, O. A. (**1992**). *Acoustics of Layered Media II: Point Sources and Bounded Beams* (Springer, Berlin).

Chew, W. C. (**1990**). *Waves and Fields in Inhomogeneous Media* (Van Nostrand Reinhold, New York).

Chew, W. C., and Kong, J. A. (**1982**). "Asymptotic approximation of waves due to a dipole on a two-layered medium," Radio Sci. **17**, 509–513.

Ciarkowski, A. (**1989**). "Uniform asymptotic expansion of an integral with a saddle point, a pole, and a branch point," Proc. R. Soc. London, Ser. A **426**, 273–286.

De Hoop, A. T., and van der Hijden, J. H. M. T. (**1984**). "Generation of acoustic waves by an impulsive point source in a fluid/solid configuration with a plane boundary," J. Acoust. Soc. Am. **75**, 1709–1715.

De Hoop, A. T., and van der Hijden, J. H. M. T. (**1985**). "Seismic waves generated by an impulsive point source in a fluid/solid configuration with a plane boundary," Geophysics **50**, 1083–1090.

Engheta, N., Papas, C. H., and Elachi, C. (**1982**). "Radiation patterns of interfacial dipole antennas," Radio Sci. **17**, 1557–1566.

Felsen, L. B., and Marcuvitz, N. (**1994**). *Radiation and Scattering of Waves* (IEEE Press, New York).

Frisk, G. V. (**1994**). *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (Prentice Hall, Englewood Cliffs, NJ).

Gallop, J. B., and Hron, F. (**1999**). "Asymptotic solutions to Cagniard's problem," Geophys. J. Int. **138**, 820–830.

Gradshteyn, I. S., and Ryzhik, I. M. (**1994**). *Table of Integrals, Series, and Products*, 5th ed. (Academic, San Diego, CA).

Gridin, D. (**1998**). "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.

Jeffrey, A., and Zwillinger, D. (**2005**). "Gradshteyn and Ryzhik Table of Integrals, Series, and Products," gr6_errata.pdf, URL http://www.mathtable.com/gr/. Last accessed 5 February 2007.

Kinsler, L. E., Frey, A. R., Coppens, A. E., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics*, 4th ed. (Wiley, New York).

Stickler, D. C. (**1976**). "Reflected and lateral waves for the Sommerfeld model," J. Acoust. Soc. Am. **60**, 1061–1070.

Wolfram, S. (**1999**). *The Mathematica Book*, 4th ed. (Cambridge University Press, New York).

Wong, R. (**2001**). *Asymptotic Approximations of Integrals* (SIAM, Philadelphia).

Detection and localization of rib detachment in thin metal and composite plates by inversion of laser Doppler vibrometry scans

Anthony J. Romano,[a] Joseph A. Bucaro, Joseph F. Vignola,[b] and Phillip B. Abraham
Naval Research Laboratory, Washington, D.C. 20375

The laboratory implementation of a fault detection and localization method based on inversion of dynamic surface displacements measured by a scanned laser Doppler vibrometer (SLDV) was investigated. The technique uses flexural wave and generalized force inversion algorithms which have previously been demonstrated using simulated noise-free vibration data generated for thick plates with a finite element model. Here these inversion algorithms to SLDV measurements made in the laboratory on a thin nickel plate and a thin carbon fiber composite plate, both having attached reinforcing ribs with intentional de-bonding of the rib/plate interface at a specific location on each structure are applied. The inverted displacement maps clearly detect and locate the detachment, whereas direct observation of the surface displacements does not. It is shown that the technique is relatively robust to the choice of frequency and to the presence of noise. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2715459]

Pages: 2667–2672

I. INTRODUCTION

Interest in infrastructure health monitoring has continued to grow over the past decade. Monitoring techniques using the dynamic mechanical, acoustic, or ultrasonic response represent one of several major research focuses aimed at providing increased capabilities for monitoring fault development in structures. These mechanical approaches include modal analysis techniques (see, e.g., Refs. 1 and 2) based on monitoring resonant frequencies and/or mode shapes associated with global structural vibration. One drawback of these modal approaches results from the fact that local changes in a structure often produce only very small changes in the global modal properties whereas unavoidable environmental changes can have a large impact. In addition, localization of the flaw is difficult using global mode responses.

The method discussed here, recently reported by Bucaro *et al.*,[3] also uses the global mechanical dynamic response; however it is able to detect *local* changes in the structural dynamics caused by the presence of a fault and thereby localize it. This method, like modally based techniques, uses measurements of surface displacement excited on a plate-like structure in response to a local, dynamic force. Various inversion techniques applied to the surface displacements serve to detect and locate the interior faults. Bucaro *et al.*[3] demonstrated these inversion algorithms using "noise-free" numerical data generated on thick, metal plates using a finite element dynamic code. In the present work, we apply two such inversion algorithms to scanned laser Doppler vibrometer (SLDV) measurements made in the laboratory on a thin

nickel plate and a thin fiber composite plate, each having attached reinforcing ribs with intentional delamination of the rib/plate bonding at a specific location on each structure. In so doing, we are able to demonstrate an actual "system" based on this approach and to examine practical issues such as noise and frequency band, both of which are important for various applications of such a technique. Further, one of the structures—the composite plate—is a material of great significance in the aircraft industry which continues to search for suitable fault-monitoring techniques which could be effective in preventing catastrophic failure of wing and stabilizer structures made from these composite materials.

In Sec. II we present by way of review the inversion algorithms. In Sec. III, we present the measured data and the results of the application of these algorithms. In Sec. IV we discuss the impact of noise and how the fault detection algorithms depend on frequency.

II. MATHEMATICAL DESCRIPTION OF THE INVERSION ALGORITHMS

We consider three techniques for using the measured surface displacement data in order to detect and locate an interior fault. The first is direct observation of the spatial distribution of the surface displacement while the other two are considered inverse techniques. One technique involves the variational inversion of the equations of motion to obtain elastic moduli maps similar to that used by Romano *et al.*[4,5] for inversion of shear wave displacement in soft, tissue-like materials; the second method is a generalized force mapping technique using a form of the inhomogeneous equations of motion.

[a]Electronic mail: anthony.romano@nrl.navy.mil
[b]Present address: Department of Mechanical Engineering, The Catholic University of America, 620 Michigan Ave., N.E., Washington, DC 20064. Electronic mail: vignola@cua.edu

A. Direct observation of surface displacement spatial maps

In some cases, the nature of the interior flaw is such that it leads to large local displacements in its neighborhood which are readily observable in the surface displacement maps themselves. In such cases, direct observation of the displacement data clearly indicates the presence and location of the internal flaw. This direct observation technique has recently been used successfully to identify fault layers in plaster walls and ceilings bearing historically invaluable frescoes in the United States Capitol Building[6] using surface displacement data obtained with a scanning laser Doppler vibrometer. We shall show that in the case of rib detachment on the plate systems studied here, the surface displacement maps do not directly reveal the flaws.

B. Variational inversion

As demonstrated by Bucaro et al.,[3] the equation of motion appropriate for a thin, isotropic plate

$$D\left(\frac{\partial^4 u_z}{\partial x^4} + 2\frac{\partial^4 u_z}{\partial x^2 \partial y^2} + \frac{\partial^4 u_z}{\partial y^4}\right) - \rho\omega^2 h u_z = f_z(x,y,\omega), \quad (1)$$

can be expressed in variational form as

$$\int_x\int_y\left(v_z D\left(\frac{\partial^4 u_z}{\partial x^4} + 2\frac{\partial^4 u_z}{\partial x^2 \partial y^2} + \frac{\partial^4 u_z}{\partial y^4}\right) - v_z\rho\omega^2 h u_z\right)dxdy$$

$$= \int_x\int_z v_z f_z(x,y,\omega)dxdy, \quad (2)$$

where h is the plate thickness, ρ the plate density, D is the flexural rigidity defined as $Eh^3/12(1-\nu^2)$ (where E is Young's modulus and ν is Poisson's ratio), f_z is an applied normal surface force, and v_z is our two-dimensional virtual function defined as $v_z=(1-\bar{x}^2)^2(1-\bar{y}^2)^2$. For a surface area element of sides L_m^x, L_m^y, centered on the global coordinates (x_m, y_m), we defined the local coordinates $(\bar{x}, \bar{y})$ as $\bar{x}=2(x-x_m)/L_m^x$, and $\bar{y}=2(y-y_m)/L_m^y$.

Integration of Eq. (2) by parts four times yields

$$D\left[\int_y\left[\left.\frac{\partial^2 v_z}{\partial x^2}\frac{\partial u_z}{\partial x}\right|_x - \left.\frac{\partial^3 v_z}{\partial x^3}u_z\right|_x + \int_x u_z\frac{\partial^4 v_z}{\partial x^4}dx\right]dy + 2\int_x\int_y u_z\frac{\partial^4 v_z}{\partial x^2 \partial y^2}dxdy \\ + \int_x\left[\left.\frac{\partial^2 v_z}{\partial y^2}\frac{\partial u_z}{\partial y}\right|_y - \left.\frac{\partial^3 v_z}{\partial y^3}u_z\right|_y + \int_x u_z\frac{\partial^4 v_z}{\partial y^4}dy\right]dx\right] - \int_x\int_y\rho\omega^2 h v_z u_z dxdy = \int_x\int_y v_z f_z dxdy. \quad (3)$$

Here, we see that the variational form has forced all but first order derivatives onto the virtual functions, thereby reducing the effects of noise in the process of calculating higher order derivatives. For simplicity, let us label the integrals multiplying the flexural rigidity, D, in Eq. (3) as I and rewrite our expression as

$$DI - \int_x\int_y\rho\omega^2 h v_z u_z dxdy = \int_x\int_y v_z f_z dxdy. \quad (4)$$

Away from any applied force (where $f_z=0$) and dividing by the quantity ρh, Eq. (4) can be rearranged to yield

$$\left(\frac{D}{\rho h}\right)_m = \frac{\int_x\int_y\omega^2 v_z u_z dxdy}{I}. \quad (5)$$

C. Generalized force mapping

As discussed in Ref. 3, the concept we termed the generalized force mapping results from a straightforward application of the variational form of the inhomogeneous equation of motion as portrayed in Eq. (4). This approach is not an "inversion" algorithm per se, but rather it provides a mapping of forces or faults through a forward calculation of Eq. (4), which was shown to be less susceptible to noise artifacts than the inversion as shown in Eq. (5).[3] In this method, we

assumed not only that we had a knowledge of the measured displacement field, $u_z(x,y,\omega)$, but also that the material parameters and the plate thickness were known and assumed to be constant with respect to space within each local voxel of interrogation. With regards to Eq. (4), we divided both sides by the quantity ρh. We saw that for a homogeneous plate away from the applied force, the two terms on the left-hand side summed identically to zero. For a plate with an inclusion or applied surface traction, we interpreted a nonzero result for the left-hand side (away from the known position of the applied force) as a generalized force G_m defined as
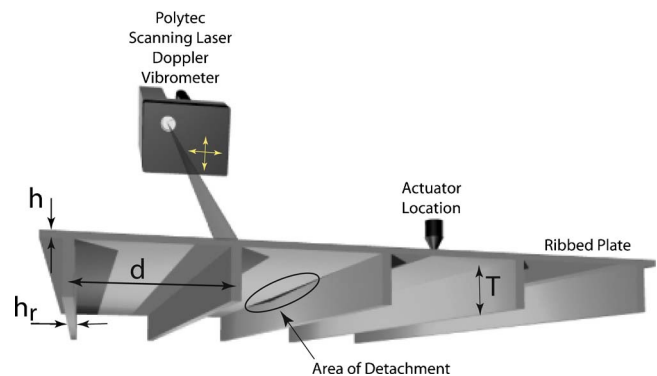
FIG. 1. (Color online) Diagram of experimental situation showing ribbed plate, actuator location, scanning laser Doppler vibrometer, and plate dimension parameters.
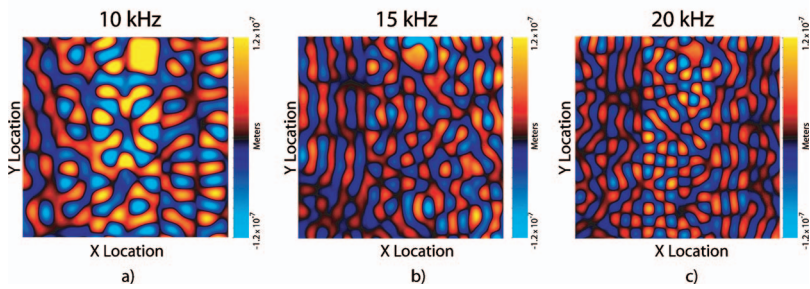
FIG. 2. Experimental results showing surface displacement for ribbed nickel plate excited at (a) 10 kHz, (b) 15 kHz, and (c) 20 kHz.

$G_m = \int_x \int_y (v_z f_z(x, y, \omega)/\rho h) dx dy$, which existed in the effected region as a consequence of the presence of the flaw. These cases may be succinctly expressed as

$$\left(\frac{DI}{\rho h}\right)_m - \int_x \int_y \omega^2 v_z u_z dx dy$$

$$= \begin{cases} 0 & \text{if the plate is homogeneous;} \\ G_m & \text{if a force or flaw is present.} \end{cases} \quad (6)$$

Therefore, as the left-hand side of Eq. (6) is calculated at each two-dimensional voxel centered at the location x_m, y_m, over the surface of the plate, nonzero values illuminate and map the spatial distributions of any faults or forces. In principle, the spatial and spectral details of the derived force could be used to further characterize the inhomogeneity although this possibility has not yet been explored. It should be mentioned that although we have not included the effects of attenuation in our models, these effects can easily be accounted for as discussed in Refs. 4 and 5, and will be the focus of future work in this area.

III. LABORATORY MEASUREMENTS AND APPLICATION OF THE INVERSION ALGORITHMS

A SLDV-based surface displacement measurement system was implemented and applied in the laboratory to two ribbed, thin plate structures each having an intentionally created short unattached rib section. One plate was fabricated from nickel and the ribs welded in place and the other from a carbon-fiber composite of the type used in airframe structures with the ribs epoxied in place. The graphite plate is manufactured by Aerospace Composite Products from multiple plies of unidirectional graphite in a cured epoxy matrix.

The general plate geometry is depicted in Fig. 1. The dimensions of the nickel plate were 30 cm×30 cm with a

thickness of $h = 1.27$ mm. There were seven ribs (of height $T = 1.1$ cm and thickness $h_r = 1.27$ mm) separated by a distance $d = 4.9$ cm. There was an intentionally unwelded segment 11.2 cm in length along the fourth rib in the center of the plate. Correspondingly, the dimensions of the carbon fiber plate were 26.9 cm×26.1 cm with a thickness of $h = 1.66$ mm. Again, there were seven ribs (of height $T = 8.36$ mm and thickness $h_r = 3.14$ mm) separated by a distance $d = 4.1$ cm. There was an intentionally unglued segment 9.0 cm in length along the fourth rib in the center of the plate.

The overall experimental arrangement is also shown in Fig. 1. The plate to be measured was placed face up on four bricks with its corners constrained by four wooden L-brackets epoxied to the bricks. Vibration was introduced into the plate structure by a Wilcoxon F-7 piezoelectric shaker mounted directly to the plate at the location indicated in Fig. 1 and driven by a linear FM chirp with an approximate bandwidth of 5–25 kHz. We chose this bandwidth such that the structural wavelengths on the plates would be on the order of the fault sizes we were attempting to illuminate. A commercially available Polytec PSV-200 scanning laser vibrometer was used to measure the out of plane (normal) surface velocity v_z which is related to the displacement by $-j\omega u_z = v_z$. Sixteen velocity records of 512 samples, collected at a sampling frequency of 64 kHz, were digitized and averaged at each spatial location. In the case of the nickel plate, there were 245×245 points across the surface with an isotropic spatial discretization of 1.24 mm; for the carbon fiber plate, data were taken at each of 259×251 points with an isotropic spatial discretization of 1.04 mm.

As a typical example, in Fig. 2 we show the measured displacements for the point actuated, ribbed nickel plate at 10, 15, and 20 kHz, i.e., over the central region of the exci-
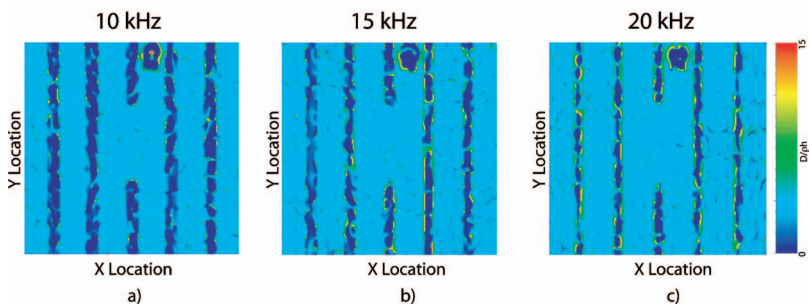
FIG. 3. Experimental results showing application of the variational inversion for the quantity $D\backslash\rho h$ for ribbed nickel plate at (a) 10 kHz, (b) 15 kHz, and (c) 20 kHz.
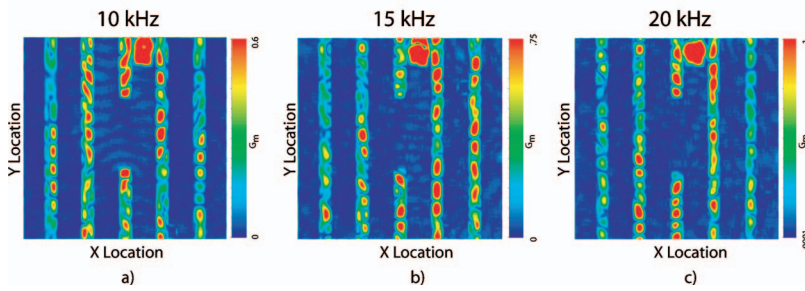
FIG. 4. Experimental results showing application of the generalized force mapping algorithm for the quantity G for ribbed nickel plate at (a) 10 kHz, (b) 15 kHz, and (c) 20 kHz.

tation band. One is not able to locate the ribs, the delamination, or the actuator unambiguously from observation of the displacement information alone. Application of the variational inversion for the quantity $D/\rho h$ using Eq. (5) is shown in Fig. 3; these images clearly illuminate these respective areas and most importantly both the ribs and where detachment occurs. To provide these images, the integrations in Eq. (5) were calculated over 16×16 data points, centered at each of 230×230 pixels, (x_m, y_m), i.e., $x_m \pm 8\Delta x$, $y_m \pm 8\Delta y$, yielding surface elements with sides $L_m^x = L_m^y = 1.984$ cm.

In the images of Fig. 3, dark blue vertical lines successfully track the locations of the hidden, interior ribs, and the associated gap in the middle of the center frame line successfully indicates where the rib is not attached. Away from the ribs or the location of the force, the average value for the quantity $D/\rho h$ is seen to be 3.5 (N m^3)/kg, a value in excellent agreement with the value of 3.55 (N m^3)/kg predicted using the following estimates for the material parameters:[7] $E = 21 \times 10^{10}$ Pa, $v = 0.31$, and $\rho = 8800$ kg/m^3.

Using the value of 3.5 (N m^3)/kg for the calibration value in the generalized force mapping algorithm as described by Eq. (6), we obtain the images shown in Fig. 4. In these images, the uniform, nonribbed areas of the plate have values very close to zero representing the homogeneous case of Eq. (6), whereas one observes nonzero values for G at the ribs and driver (thus exposing them) which represents the inhomogeneous case in Eq. (6).

In Fig. 5 we show the measured displacements for the point actuated, ribbed carbon fiber composite plate at 10, 15, and 20 kHz, respectively. As in the case of the nickel plate, one is not able to locate the ribs, the delamination, or the actuator position. Application of the variational inversion from Eq. (5), however, clearly illuminates these respective areas as shown in Fig. 6 for the three frequencies described above. To provide this image, the integrations in Eq. (5) were

calculated over 16×16 data points, centered at each of 244×236 pixels (x_m, y_m), i.e., $x_m \pm 8\Delta x$, $y_m \pm 8\Delta y$, yielding surface elements with sides $L_m^x = L_m^y = 1.66$ cm. In this image, dark blue vertical lines once again successfully track the locations of the hidden, interior ribs, and the associated gap in the middle of the center rib line successfully indicates where detachment occurs. Away from the ribs or force location, the average value for the quantity $D/\rho h$ is seen to be 6.5 (N m^3)/kg, however, in this case, we do not have good estimates of the particular material parameters needed to predict $D/\rho h$ as we did for the nickel plate.

Using the value 6.5 (N m^3)/kg for $D/\rho h$ in the generalized force mapping algorithm as described in Eq. (6), we obtain the image shown in Fig. 7. In this image, the nonribbed areas of the plate are very close to zero once again, representing the homogeneous case of Eq. (6), while the ribs and the driver are clearly illuminated portraying a nonzero value for G, which represents the inhomogeneous case in Eq. (6).

IV. DISCUSSION

The results presented in the previous section demonstrate and establish the feasibility of an internal fault monitoring method for plate-like structures using inversion of SLDV measured surface displacement. Our focus here is on monitoring rib detachment in thin plate structures, and both the flexural inversion algorithm as described by Eqs. (3)–(5) and the generalized force algorithm as described by Eq. (6) are seen to work well in detecting and locating rib detachment. These results, together with those in the simulation studies of Ref. 3, indicate that the technique reported here would be effective for other internal fault conditions as well, such as inclusions, voids, and local material property changes. Regarding this last application, as the flexural in-
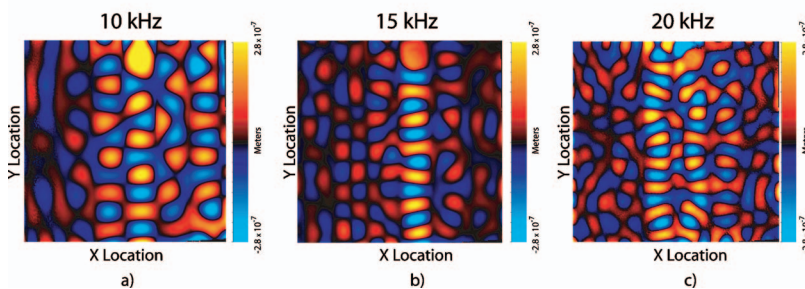
FIG. 5. Experimental results showing surface displacement for ribbed carbon-fiber composite plate excited at (a) 10 kHz, (b) 15 kHz, and (c) 20 kHz.

Romano *et al.*: Detection and localization of rib detachment

Carbon Fiber Variational Inversion

10 kHz   15 kHz   20 kHz

a)   b)   c)
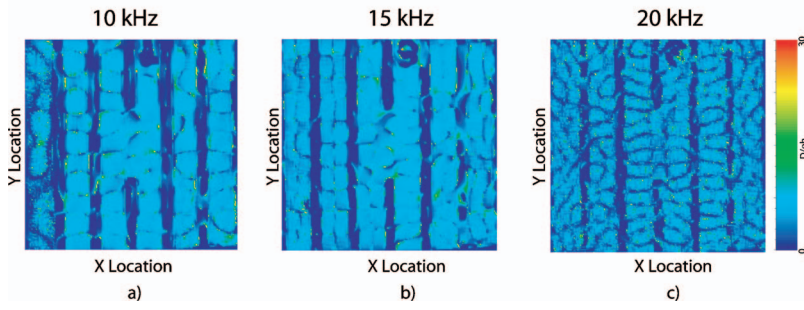
FIG. 6. Experimental results showing application of the variational inversion for the quantity $D\backslash\rho h\}$ for ribbed carbon-fiber composite plate at (a) 10 kHz, (b) 15 kHz, and (c) 20 kHz.

version algorithm provides a direct spatial mapping of the effective plate stiffness parameter $D/\rho h$, it is particularly suited for quantitative monitoring of local material property variations.

Even though our focus here has centered on fault morphologies whose sizes fall in the centimeter range, this structural acoustic method does not appear to be limited to any particular length scale. Indeed, the inversion algorithms themselves are independent of scale. What is required is (1) that the structure be mechanically excited at frequencies, ω, whose *structural* wavelengths, λ_{struc}, are short enough to be measurable by the SLDV in the presence of noise over the available scan distance and (2) that the dynamic surface displacements be mapped with a spatial sampling an order of magnitude smaller than the fault size, a consequence of the need to sufficiently resolve the position and extent of the flaw.

The first condition is necessary to prevent ill-conditioned behavior from application of the inversion algorithms. Straightforward consideration of the spatial derivative shows that measurability of the structural wavelength requires that $(k_{struc}\Delta x)^{-1} \leq S/N$, where S is the displacement signal and N the total background noise. We arrived at this expression from the following simple reasoning: Consider a wave with wavenumber k_{struc} which we want to reconstruct as we scan in the presence of noise, N. For a change in position, Δx, the largest signal change, $Sk_{struc}\Delta x$, will be "measureable" when it is at least equal to the noise level, N. This yields the condition

$$\frac{S}{N} = (k_{struc}\Delta x)^{-1}. \tag{7}$$

Therefore, if $(k_{struc}\Delta x)^{-1}$ is less than or equal to S/N, then it will be possible to map the signal of wavenumber k_{struc}.

As an example, consider flexural wave inversion and the case in which the displacement signal to noise level is independent of frequency. In this case[8]

$$k_{struc} = \frac{2\pi}{\lambda_{flex}} = \left(\frac{D}{\rho h}\right)^{-1/4}\sqrt{\omega} = C\sqrt{\omega} \tag{8}$$

with $C = \{D/\rho h\}^{-1/4}$ resulting in the condition $(S/N) \geq \{C\sqrt{\omega}\Delta x\}^{-1}$. This shows that at the low end of the frequency band one would encounter a low frequency limit below which stable inversions would not be possible. It also can be seen from this condition and Eq. (8) that the higher the S/N, the longer the wavelength can be for a given spatial discretization. At higher frequencies, one can encounter a high frequency limit due to material damping effects which typically increase with frequency eventually leading to an unacceptable drop in S/N.

A scanning laser Doppler vibrometer differs from other metrology tools in that the noise level varies for each measure in a scan. It is the speckle[9] character of the backscattered light which produces a unique distribution of intensity and phase in the optical collection aperture that is responsible for these variations. The SLDV reacquires data at locations where the noise exceeds a predetermined threshold. These reacquisitions are collected at spatial locations that may differ from the original measurement location by a small fraction of the spatial sampling interval to change the backscattered speckle and reduce the noise. The variations in noise range over more than an order of magnitude and have a mean root mean square value of approximate 10^{-6} m/s over the measurement band. The surface velocities on both the carbon fiber and nickel structures exceeded 10^{-2} m/s in the regions adjacent to the drive points with antinode velocities on the carbon fiber structure being about twice those

Carbon Fiber Generalized Force Mapping

10 kHz   15 kHz   20 kHz

a)   b)   c)
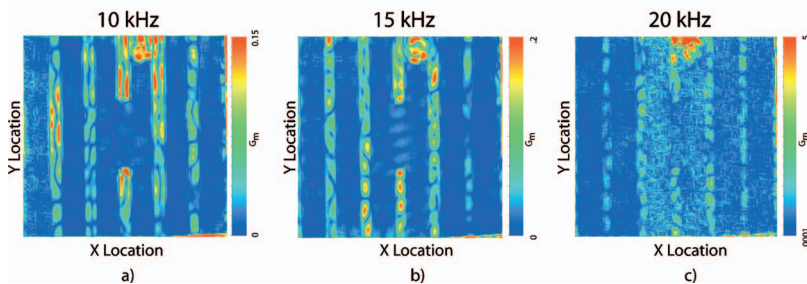
FIG. 7. Experimental results showing application of the generalized force mapping algorithm for the quantity G for the ribbed carbon-fiber composite plate at (a) 10 kHz, (b) 15 kHz, and (c) 20 kHz.

seen on the nickel structure. Antinode peak velocities fall off with distance in the direction across the rib by 25 dB whereas they diminish very little with distance in the direction parallel to the ribs. This implies that the highest values for signal to noise ratio exceed 80 dB. The lowest values of the surface velocity are found adjacent to the attached ribs and in particular the two ribs at the outer edges. These surface velocities are 30–50 dB lower than the peak velocities. At these locations the signal to noise ratio is not less than 30 dB.

In the carbon composite case, some artifacts appear in the two inversion results at the two upper frequencies (15 and 20 kHz) leading to a decrease in performance compared to the 10 kHz case for both the flexural inversion and generalized force algorithms (Figs. 6 and 7). One cause for these artifacts that comes to mind is a decrease in signal to noise ratio at the higher frequencies in the carbon fiber composite plate SLDV displacement measurements. This might be expected due to the higher levels of vibration damping for the composite material and because those levels are also expected to increase with frequency. However, the S/N levels in the measured displacement scans appear comparable for the nickel and carbon composite cases (see Figs. 2 and 5), although there is a greater tendency in the latter case for the vibration levels to be higher in the center regions of the plate than toward the two edges. A second possibility is that the isotropic equation of motion for a thin plate may not be appropriate for the carbon fiber composite plate which is actually anisotropic and may then lead to anomalies in the inversion and generalized force mappings. However, further work is required to resolve this issue.

V. CONCLUDING REMARKS

We have reviewed the formulation of several fault detection algorithms and have demonstrated their practical application in the laboratory using the SLDV technique. In one algorithm, an inversion of the measured surface displacements provided a mapping of the flexural rigidity of a plate scaled by the density and thickness in a local fashion, while the other algorithm was based on a generalized force mapping which illuminated the locations of faults or forces present on a plate, also in a local fashion. These algorithms were applied to two reinforced thin plate structures: a ribbed nickel plate and a ribbed carbon fiber composite plate, both of which had an intentional area of detachment in their rib

structures. The results indicate that the algorithms are robust in such applications, and that together they are effective in both providing material parameter maps and locating faults and forces.

The structural acoustic methodology discussed here can offer a reliable, practical approach to mechanical health monitoring in a variety of structures. It combines the suffusive nature of broadband structural vibration with the power of carefully designed inversion algorithms in order to detect and *locate* various faults within plate-like structures. For fault monitoring applications, plate results are more general than one might first think in that many structures can be addressed at some level as a collection of plate-like elements. The methodology should have general application, from large scale structures such as ships and aircraft down to micro- and nanostructures. This approach is relatively new, and as the sophistication of the inversion algorithms grows, the fault detection performance and its identification ability should continue to advance.

[1] M. L. Wang, F. L. Xu, and G. M. Lloyd, "A systematic numerical analysis of the damage index method used for bridge diagnostics," Proc. SPIE **3988**, 154–164 (2000).

[2] J. T. Kim and N. Stubbs, "Model-uncertainty impact and damage-detection accuracy in plate girder," J. Struct. Eng. **121**, 1409–1417 (1995).

[3] J. A. Bucaro, A. J. Romano, P. Abraham, and S. Dey, "Detection and localization of inclusions in plates using inversion of point actuated surface displacements," J. Acoust. Soc. Am. **115**, 201–206 (2004).

[4] A. J. Romano, J. J. Shirron, and J. A. Bucaro, "On the non-invasive determination of material parameters from a knowledge of elastic displacements: theory and numerical simulation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 751–759 (1998).

[5] A. J. Romano, J. A. Bucaro, R. L. Ehman, and J. J. Shirron, "Evaluation of a material parameter extraction algorithm using MRI-based displacement measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1575–1581 (2000).

[6] J. F. Vignola, J. A. Bucaro, B. R. Lemon, G. W. Adams, A. J. Kurdila, B. Marchetti, E. Esposito, E. Tomasini, H. J. Simpson, and B. H. Houston, "Locating faults in wall paintings at the U.S. capitol by shaker-based laser vibrometry," APT Bulletin, the J. Preservation. Tech., 25–33 (2005).

[7] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982), p. 461.

[8] G. Williams, *Fourier Acoustics* (Academic Press, San Diego, 1999), pp. 56–60.

[9] S. Rothberg, "Numerical simulation of speckle noise in laser vibrometry," Appl. Opt. **45**, 4523–4533 (2006).

Noise in the operating rooms of Johns Hopkins Hospital

Jonathan M. Kracht, Ilene J. Busch-Vishniac,[a] and James E. West

Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218

Very little reliable information exists on the sound levels present in an operating room environment. To remedy this situation, sound pressure levels of the operating rooms in Johns Hopkins Hospital were monitored before, during, and after operations. The data were analyzed to determine background sound levels, average equivalent sound levels L_{eq}, frequency distribution, and peak sound pressure levels L_{peak}. Each surgery was matched to the period of noise it produced permitting the association of sound levels with particular types of surgeries and the determination of various sound measures for classes of surgery (e.g., orthopedic, neurological, etc.). Averaging over many surgeries, orthopedic surgery was found to have the highest L_{eq} at approximately 66 dB(A). Neurosurgery, urology, cardiology, and gastrointestinal surgery followed closely, ranging from 62 to 65 dB(A). By considering the L_{peak} along with the L_{eq} values, a pattern emerges for the various surgical divisions. Gastrointestinal and thoracic surgery are relatively quiet among the surgical divisions. Neurosurgery and orthopedics have sustained high sound levels. Cardiology surgery has a more moderate average sound level but includes brief periods of extremely high peak sound levels. For neurosurgery and orthopedic surgery, peak levels exceeded 100 dB over 40% of the time. The highest peak levels routinely seen during surgery were well in excess of 120 dB. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2714921]

Pages: 2673–2680

I. INTRODUCTION

In spite of the fact that noise is a top complaint of patients and staff in hospitals,[1,2] there has been relatively little written which characterizes hospital noise and outlines control strategies. (A recent paper by Busch-Vishniac *et al.*[3] presents an extensive review of the hospital noise literature.)

Within hospitals, there is growing concern about noise in the operating theaters, but the literature on this topic is generally thin and often conflicting. The literature varies from editorials railing against noise in the operating room (OR)[4,5] to studies of the noise produced by tools used primarily in orthopedic surgery[6–8] and to reports of the sound pressure levels produced by a particular surgery. While rare, there are also a few reports of the effects of OR noise on performance and hearing. Moorthy *et al.*,[9] for instance, studied the ability of surgeons to place three laparoscopic sutures on a suture pad in quiet and noisy conditions using background operating room noise and music. They found no significant differences in performance attributable to the presence of sound. Willett *et al.*[10] found, however, that about half of the orthopedic surgeons in their study showed significant noise-induced hearing loss, which they presumed was attributable to their exposure to the loud noises of orthopedic surgical equipment such as bone saws and drills.

Several studies have measured the noise levels for specific surgeries. Bovenzi and Collareta[11] measured an equivalent sound pressure level (L_{eq}) of 64.8 dB(A) during a 4 h aorto-coronary bypass graft. Shankar *et al.*[12] noted four surgeries in different fields of medicine. They reported an L_{eq} of 58 dB(A) for a tonsillectomy in an otolaryngology OR, 60 dB(A) for a McMurray's osteotomy in the orthopedic OR, 54–56 dB(A) for an exploratory laparotomy with hysterectomy in a gynecology OR, and 60 dB(A) for a herniotomy in the surgery OR. Hodge and Thompson[13] reported a L_{eq} of 51 dB(A) for a radical neck dissection. Focusing primarily on orthopedics, Love[14] observed an average of 79 dB(A) for total hip replacement surgery and 81 dB(A) for total knee replacement.

Murthy *et al.*[15] conducted a study of the noise in the operating room, noting its effect on anesthesiologists who work within. Instead of reporting levels for individual surgeries, they gave levels for various divisions of surgery. Neurological operating rooms had the highest L_{eq} associated with them at 78.2 dB(A). Emergency surgery, orthopedics and general surgery followed closely at 77.2, 76, and 74.6 dB(A), respectively. Cardiology was measured to be the most quiet at 69 dB(A).

Additionally, there have been some reports in the literature of the background noise found in ORs. Nott and West,[16] for instance, state that background noise in the OR registered an L_{eq} of between 50 and 60 dB(A). Similarly, Shankar *et al.*[12] found a background L_{eq} for otolaryngology, gynecology, and general surgery of approximately 48 dB(A) while orthopedic surgery is at 58 dB(A). Two more rigorous studies provide not only the background L_{eq} but also the frequency spectra associated with it. Bovenzi and Collareta[11] acquired measurements in an operating room when the ventilation system is on (L_{eq} of 63 dB(A)) and off (L_{eq} of 35 dB(A)). Wallace *et al.*[17] measured the background noise in 24 operating rooms in hospitals located throughout Maryland (but not including Johns Hopkins Hospital). They specifically chose times which would minimize background

noise and measured with all equipment turned off except the air handling system. Under these conditions they found background levels ranging from 46 to 57 dB(A) with an average of 52 dB(A).

Unfortunately, it is very difficult to build up a general soundscape of the noise in an OR by assembling information taken from these various articles. Because the results cited above correspond to different locations and different types of surgery, it is not surprising that comparisons are problematic at best. However, attempts to compare and contrast the data also are hindered by the lack of information about specific measurement techniques used. The papers cited above, without exception, do not reveal the sound level meter time setting (fast, slow, or impulse) and many of the articles do not provide the period of time averaging for the L_{eq}.

In this paper, we aim to provide a comprehensive description of the soundscape of the ORs at a particular hospital, Johns Hopkins Hospital in Baltimore, Maryland. Results include background sound levels in octave bands for each OR, and typical L_{eq} and peak levels for particular surgeries. This rich data set permits a comparison of the sound levels for different categories of surgery. Further, as the ORs are spread over two sets of buildings—one old and one new—we are able to comment on the effect of new construction on OR noise. Johns Hopkins Hospital is a large hospital in an urban setting. It has been ranked the top hospital (by *US News and World Report*) for the last 16 years.

II. METHODS

Sound levels were measured in operating rooms throughout Johns Hopkins Hospital (JHH) to quantify the acoustical events during surgical procedures. Results are presented either in dB(A) re 20 μPa or, in the case of peak levels, in unweighted dB re 20 μPa. Each operating suite in the Weinberg, Meyer, Blalock, and Carnegie buildings was studied. Meyer, Blalock and Carnegie are old buildings in JHH while Weinberg was opened in 2000. In total, 38 rooms were measured: 16 each in Blalock/Carnegie and Weinberg and six in Meyer. The ORs vary in size from 272 to 733 ft^2 with an average footprint of 482 ft.^{2} All of the rooms have hard surfaces and furnishings with no material added for sound absorption, so all have long reverberation times. We were not able to measure reverberation time during this study because we were prohibited from disturbing the full surgical schedule.

Operations in all branches of medicine were examined including neurology, cardiology, orthopedics, urology, and plastic surgery. Surgeries performed on both adult and pediatric patients were monitored. Caution was taken to preserve the normal situation for surgery. Surgeons and staff were generally not aware of noise monitoring and no attempt was made to control conversation or the playing of music during surgery.

A Larson Davis System 824 sound level meter, serving as both a precision sound level meter and real-time frequency analyzer, was used to collect the data. This device was placed inside each room for a 24 h period. The data were stored on the sound level meter and then downloaded onto a laptop for further analysis. Readings were collected during weekdays only so as to guarantee surgical action within the room. In each trial, the sound level meter was positioned so that it did not interfere with the medical operations occurring and it was outside of the sterile field. Often, the instrument was placed on top of the fire extinguisher box in a corner of the theater. While such a location is often problematic, the ORs were generally so reverberant that we did not observe significant sound variations from location to location in each room, including along the walls and in the corners. The sound level meter was placed inside a thin plastic bag to prevent contamination and a hole provided for power connection. A paper taped to the bag explained that the instrument was present for a study although noise was not mentioned. Other than this notice, the medical staff were not informed of the ongoing study in order to avoid changes in operation protocols that might affect the noise.

A-weighted L_{eq} and flat frequency analysis in octave bands from 16 Hz to 16 kHz were obtained. Peak sound pressure levels (unweighted), L_{peak}, were also found. The rooms were analyzed using both 5 and 30 min averaging intervals for the L_{eq}. Subsequently, representative rooms were reexamined using more refined time periods of 30 s and 5 min. For all A-weighted measurements, the sound level meter slow setting was used. The unweighted L_{peak} measurements were true peak measurements based on instantaneous sound levels and limited only by the analog-to-digital converter speed (operating at about 20 μs between samples).

Once all of the data were collected, surgical logs were examined to identify operations occurring within the sampled intervals. Appropriate permission was obtained from the hospital administration to view these records. Each surgical log provides information on the nature and duration of a surgery. The surgical log notes the time of occurrence of events such as the patient entering the operating room, the first incision being made, and the patient being taken out of the room. These logs are normally quite accurate and based on the clock in each OR. We synchronized our sound level meter to this clock as well. The total surgery time was assumed to be the time from the first incision to the patient being removed. We chose the initial incision time rather than the time the patient was wheeled into the OR in order to focus specifically on the noise during surgery rather than during preparation. The surgery log also provides a detailed description of the nature of each procedure and the division of surgery under which it falls. Though available on these logs, the names of the hospital staff including the surgeon, anesthesiologist, nurses and staff as well as the patient were expunged from our record in order to preserve privacy.

Using the data obtained from the sound level meter and the surgery times, sound levels for each surgery were calculated. An A-weighted L_{eq} and octave band levels were calculated by averaging over the time of the particular surgery. Background sound levels for each operating room were obtained by selecting a point in the time history for each OR at which there was no surgery according to the logs, and which did not seem to have variable sound levels suggestive of
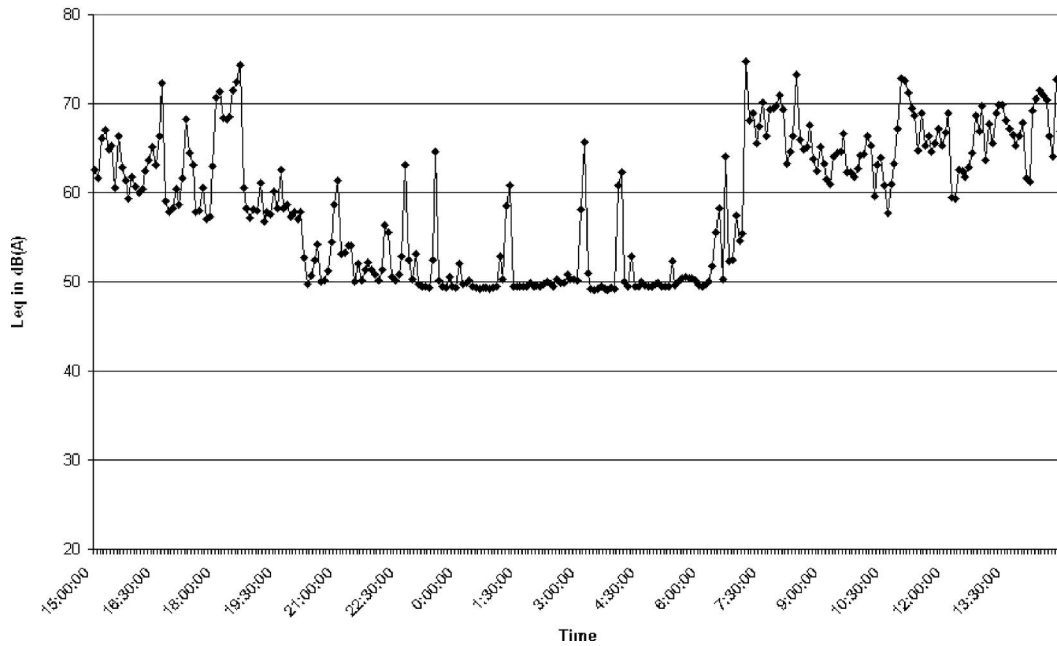
FIG. 1. A-weighted equivalent sound pressure level measured in Blalock 16 at Johns Hopkins Hospital over a 24 h period.

room cleaning or other activity. L_{peak} values were not averaged, but instead recorded for presentation in terms of fraction of the operating time various levels were exceeded.

III. RESULTS

Figure 1 shows a typical trace of A-weighted L_{eq} versus time for an operating room (Blalock 16). In this figure, L_{eq} values were obtained using a 5 min averaging time. For the period shown in this figure, there were three surgeries: pediatric plastic surgery from 15:47 to 16:38 on the first day, and pediatric urology surgeries from 8:29 to 10:59 and 12:11 to 12:50 on the second day. The times given here start with the first incision and end with the patient being removed from the OR. From Fig. 1 it is clear that there is significant noise as well during preparation of the OR for surgery and clean up after surgery. Note that the period of surgery is clearly distinguishable from the nonsurgical period although there are several brief noisy periods in the nonsurgical time and significant variability in the sound level during the surgery periods. Because we do not have sound recordings, we cannot identify the causes of the peaks in the quiet period.

The results shown in Fig. 1 make it possible to determine the background sound level as well as the average sound level for a particular surgery. For instance, one could determine the background noise level in Blalock 16 to be roughly 50 dB(A) as shown by the lowest sustained levels seen in the nonsurgical period. The high sound peaks during this period cannot be explained with the information we collected. In the results presented below, we show background sound levels as a function of location and frequency.

Figure 2 shows the background L_{eq} in each OR. There is considerable variation, about 20 dB(A), in the background levels shown. There is no striking pattern seen in comparing buildings, but Blalock is slightly less noisy than Meyer and

Weinberg. This is quite interesting since Blalock and Meyer are old buildings while Weinberg is a new building (opened in 2000) that was constructed with noise in mind and an aim at a quieter building. However, Meyer and Weinberg are closer to main entrances to JHH than Blalock so the distinction might simply reflect foot-traffic. The results shown in Fig. 2 compare well with those of Bovenzi[11] when the ventilation system was turned on. They do not compare well with those of Wallace,[17] who generally found much lower and spectrally flatter levels.

Figure 3 shows the average background noise octave band spectra for the ORs in each of the three buildings. Note that there is a great deal of similarity in the spectra for the three buildings. The shape is common to that observed in many buildings, with high levels at the low frequency end (largely attributable to air handling) and a steady drop as the frequency increases. Overall, Meyer is somewhat more noisy at the very low frequency end and in the center of the speech range. Weinberg is noisier at the upper end of the spectrum. We have not determined the cause of these variations.

Table I and Fig. 4 summarize the results for surgical noise levels presented by category of surgery (e.g., neurology, gynecology, etc.). Each result thus represents an average over multiple surgeries, typically in more than one OR. The categories in the table are ordered by increasing average L_{eq}. This L_{eq} average was calculated by averaging the results for each surgery independent of surgical time. Results in Table I are presented accurate to 0.5 dB(A).

Overall, the surgeries monitored span a range of L_{eq} from 53.0 to 70.5 dB(A), and the averages by category range from 58.0 to 67.0 dB(A). This is a tighter distribution of levels than anticipated and suggests that the prior literature might be overstating differences between types of surgeries as a result of the lack of a consistent measuring system (in-
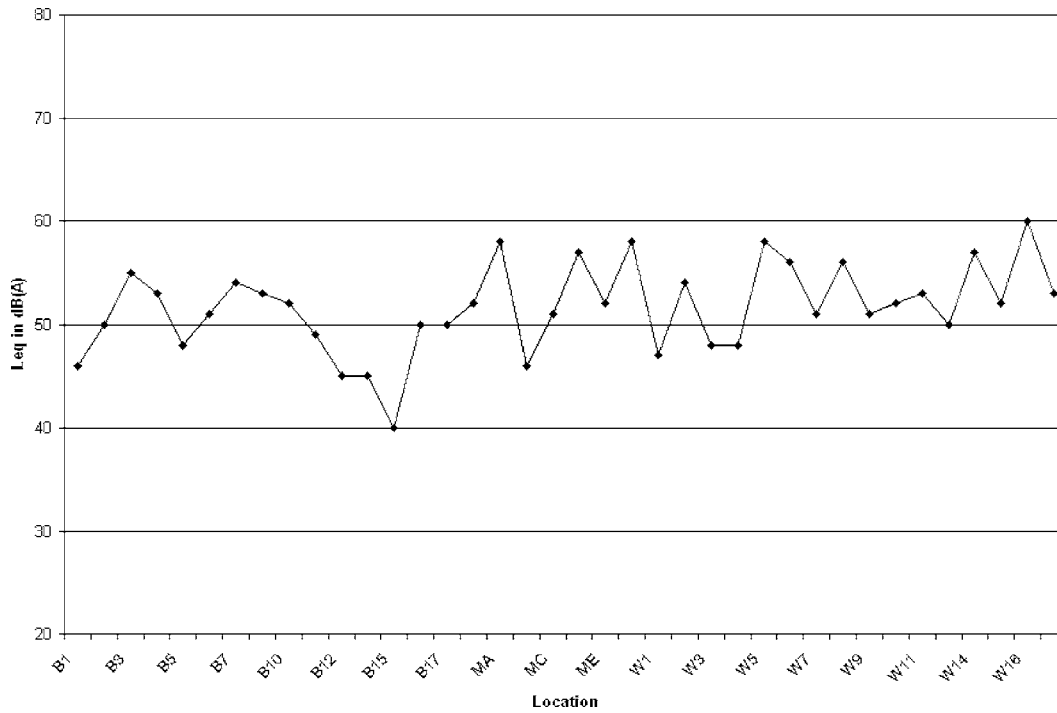
FIG. 2. A-weighted equivalent background sound pressure level of each of the operating rooms. B indicates Blalock, M Meyer, and W Weinberg.

cluding sound level meter settings). Further, it is clear from Table I that there is no discernable pattern that distinguishes the noisiest from the least noisy surgeries.

Figure 4 shows the octave band spectra by category for the surgeries monitored, averaged over all surgeries in that category without time weighting for surgical duration. Only surgery categories with more than three surgeries are shown in this figure. Note that the spectra show remarkable similarity in shape, with a broad low frequency range of high sound

amplitude, and then a decay with frequency. The broad relatively flat range includes the speech band of the 250–4000 Hz octave bands. In any given octave band, the levels span no more than 10 dB(A) across these major surgical categories. By comparing Figs. 3 and 4 we can estimate octave band signal to noise ratios for surgical noise relative to background noise in the OR. The signal to noise ratio thus computed is positive at all octave bands above 125 Hz. In the octave bands from 250 Hz to 16 kHz the signal to noise
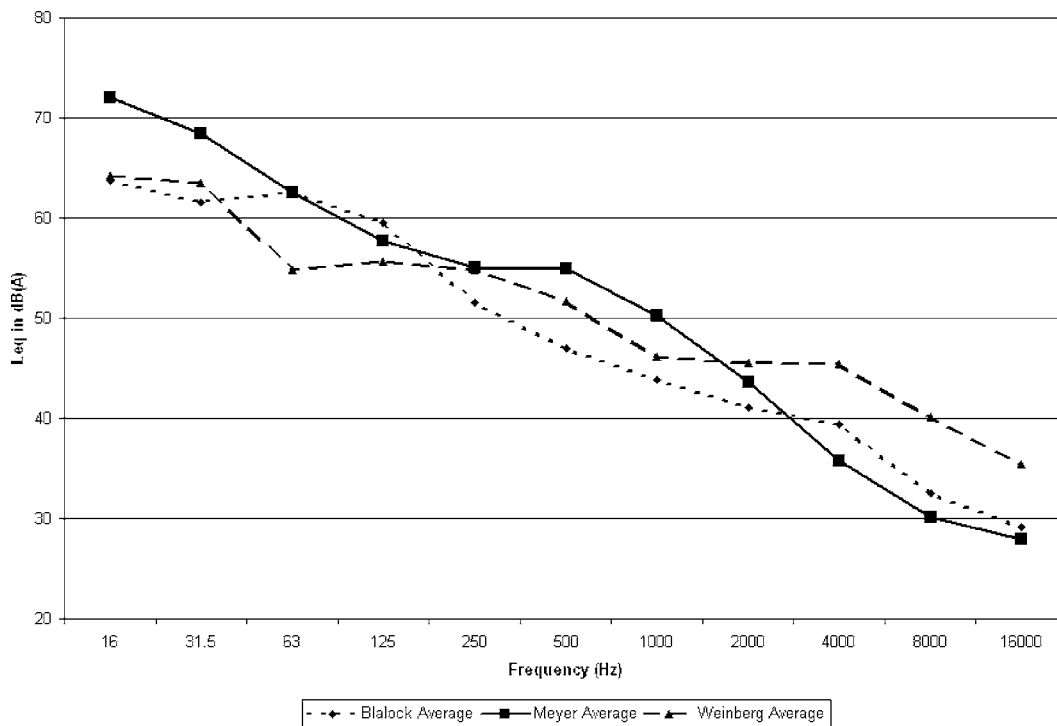
FIG. 3. Average background sound pressure level in octave bands for operating rooms in each of the three hospital buildings.

Kracht *et al.*: Operating room noise

TABLE I. Average L_{eq} and range for surgeries by category.

Division	Number of Surgeries	L_{eq} (dB(A))	dB(A) Range	Time range (min.)
Pediatric orthopedic	2	58.0	57.0–59.0	66–84
Thoracic	4	63.0	61.5–63.5	59–240
Gastrointestinal	9	63.0	55.0–69.0	39–293
Cardiology	10	63.5	57.0–68.0	40–319
Urology	11	63.5	55.5–67.0	31–157
Pediatric urology	2	64.0	64.0	39–150
Neurosurgery	8	64.5	60.0–67.5	74–510
Otolaryngology	4	65.0	53.0–66.5	36–76
Pediatric plastic	4	65.0	62.0–68.5	51–117
Orthopedic	19	66.5	56.5–70.5	19–37
Plastic	3	67.0	59.0–69.0	35–548

level varies from 3 to 24 dB(A), and averages 11 dB(A). Thus, the spectra shown in Fig. 4 represent surgical noise above 125 Hz rather than background noise.

At the high end of the spectrum, orthopedic surgery has the highest level although it is roughly average in the low- and mid frequency range. At the low frequency end of the spectrum, neurosurgery has the highest level, and it is among the highest through the entire spectral range. Pediatric plastic surgery is the least energetic in the low frequency end of the spectrum (but not least energetic at frequencies above that), and cardiology surgery was the least energetic in the high frequency end (but not the low).

Table I and Fig. 4 focus on L_{eq} values, but another metric of noise is the peak level. By examining the peak level, one can determine the frequency and intensity of acoustical events that produce very loud noise. The effect of noise peaks on hearing loss is not well understood. Berglund et al.[18] give a complete summary of work on hearing loss due to impulsive sounds, most of which comes from studies

of gunfire. This body of work served as the basis for the U.S. legal limit of 140 dB for impulsive sound exposure of workers.

Figure 5 shows the fraction of the time during surgery that the unweighted L_{peak} exceeded 90, 95, 100, and 105 dB by surgical category, where the order of divisional presentation is the same as used in Table I. The values in Fig. 5 were determined from the unweighted L_{peak} values recorded every 5 min. While a 5 min window is rather long for a peak level study, we were limited to this window because of our desire to gather continuous 24 h data and the limited memory available on the equipment. The peak information is very interesting because it provides a window on the *peakiness* of the sound in an OR during surgery.

There are two ways to consider the data in Fig. 5, both of which provide interesting information. First, it is useful to review the fraction of the time L_{peak} exceeds 95 dB in order to get a sense of how often loud acoustic events occur. We choose to examine exceedance of 95 dB rather than 90 dB simply because all of the divisions have peak levels exceeding 90 dB almost all of the time. Consideration of the exceedance of 95 dB yields information which is more sensitive to surgical division. From Fig. 5, neurosurgery leads with L_{peak} exceeding 95 dB over 78% of the time. It is followed in descending order by orthopedics, otolaryngology, urology, cardiology, and pediatric plastic surgery, all of which exhibited peak levels over 95 dB the majority of the time during surgery.

Second, by considering the highest peak levels, those over 105 dB, we can get a sense of which surgical types include very loud sound events. From the OR noise data, among the surgery categories with at least three observed surgeries, otolaryngology has the most frequent very loud noise peaks, followed in order by neurosurgery, orthopedics, pediatric plastic surgery, and urology. Figure 6 provides ad-
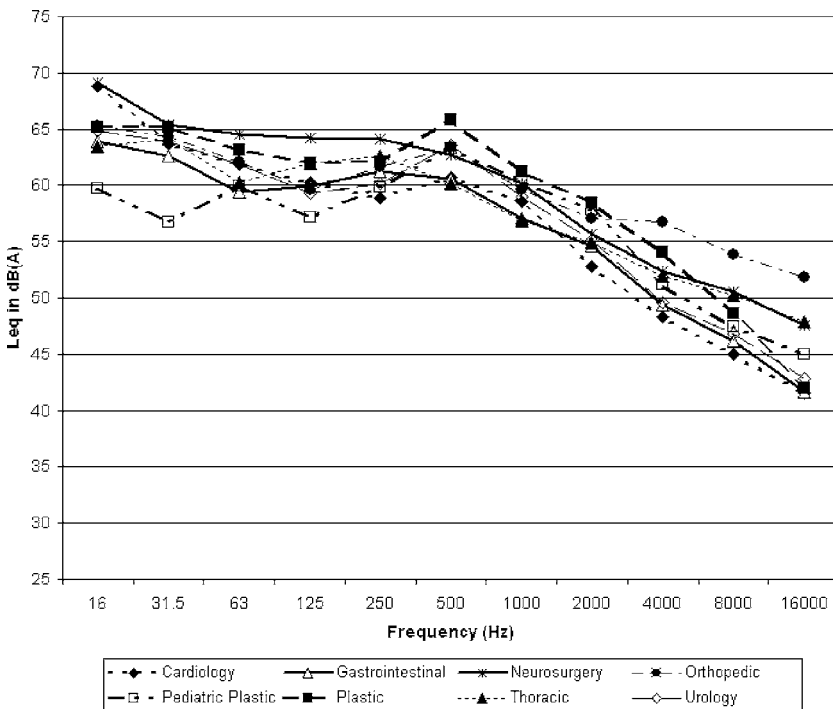
FIG. 4. Average sound pressure level in octave bands by category for surgeries.
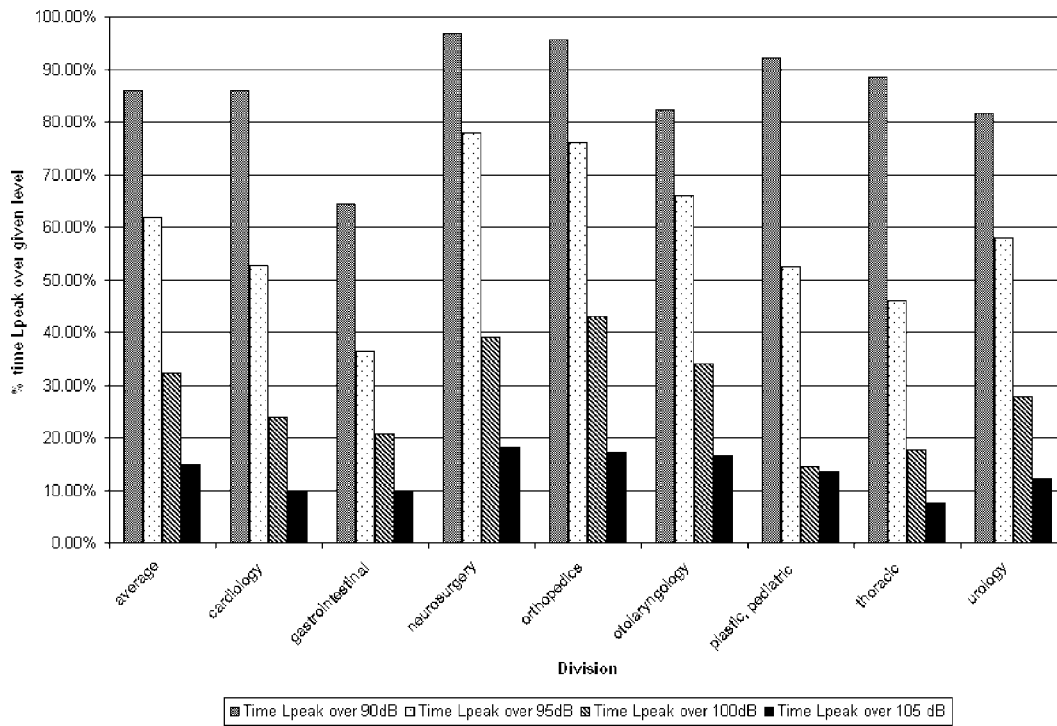
FIG. 5. Fraction of time L_{peak} exceeds 90, 95, 100, and 105 dB (unweighted) by category of surgery.

ditional information by showing the maximum value of L_{peak} obtained for each surgery. Figure 6 shows the logarithmic average of the maximum L_{peak} by division and the full range of values found within each division of surgery. This figure only includes divisions for which we have more than three surgeries monitored. From Fig. 6 we see that pediatric plastic, cardiology, and urology surgical divisions have the high-

est average maximum L_{peak} with some events recorded at over 120 dB. Gastrointestinal and thoracic surgical divisions have the lowest maximum L_{peak}.

By combining the results of Table I and Figs. 4–6 one gains a sense of the sound scape in ORs for various types of surgeries. While all of the surgeries are fairly noisy, thoracic and gastrointestinal surgeries show a pattern of less noise by
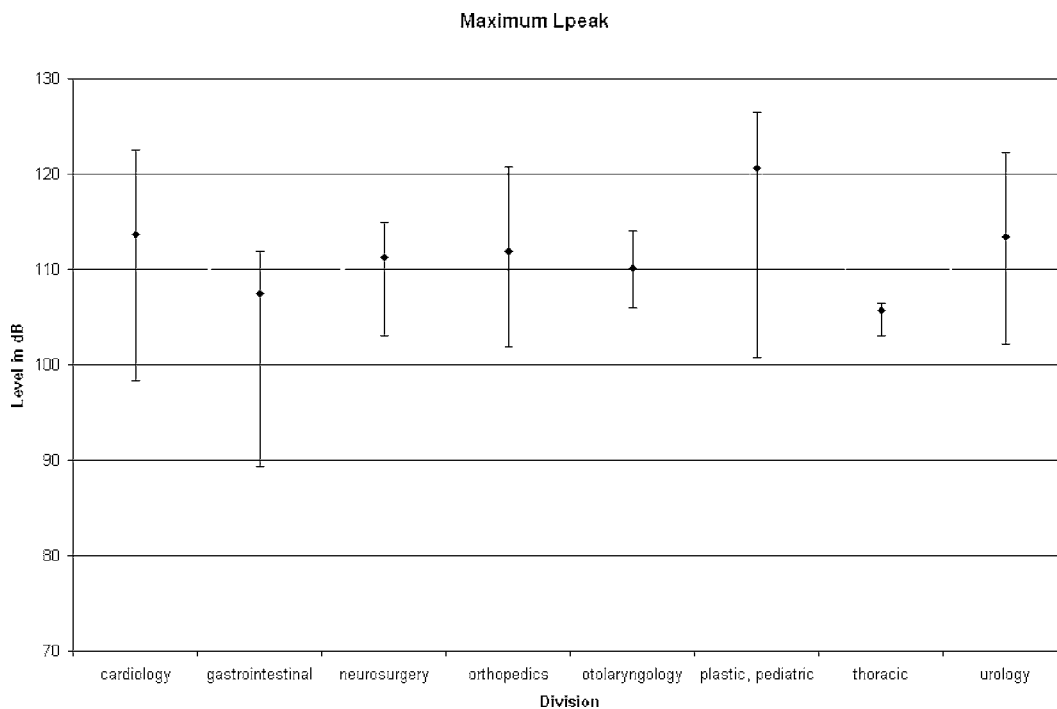
FIG. 6. Maximum L_{peak} for each surgery monitored. Results shown as dB average in surgical divisions (data point) and full range.

all metrics we monitored. They have relatively low L_{eq}, do not show many peaks exceeding 95 dB, and do not have high, brief sound levels as demonstrated by a large maximum L_{peak}. At the opposite extreme is orthopedic surgery which exhibits a relatively high L_{eq}, a L_{peak} above 95 dB most of the time, and a medium high range for the maximum L_{peak} during surgery. Cardiology and pediatric plastic surgery show a few brief but very high acoustic events. By contrast, neurosurgery is relatively intense (high L_{eq}) but has a relatively low maximum L_{peak}, so it seems to be described well as a sustained noisy environment.

Although there are certainly noise patterns which distinguish the different divisions of surgery from one another, there does not seem to be a standard acoustic signature for a given surgery. Our monitored surgeries include multiple examples of some particular surgeries, but these surgeries show marked acoustic variation, even when performed by the same surgeon on the same day. For instance, we captured data on four radical prostatectomies in two different operating rooms. For these surgeries, the L_{eq} during surgery varied from 55.6 to 64.8 dB(A). The highest L_{peak} values spanned 103.6–112.8 dB, and the fraction of time the L_{peak} exceeded 90, 95, 100, and 105 dB varied greatly.

IV. COMMENTS

The results presented here demonstrate that surgery is a noisy business. Sound pressure levels at Johns Hopkins Hospital were found to average between 55 and 70 dB(A) with significant sound peaks (some of great intensity) during surgical procedures. The concerns raised by high sound pressure levels in ORs are twofold: the potential for hearing loss, and the disruption to clear speech communication.

The sustained sound pressure levels found in the OR are not sufficiently high to cause significant hearing loss. However, the presence of high sound pressure peaks is a cause for concern. The majority of the surgical divisions we monitored experienced sound pressure peaks of at least 110 dB and levels over 120 dB were not uncommon. It is certainly possible that these very high sound pressure peaks could negatively impact the hearing of surgical staff over time. Thus, we recommend that attention be paid to identifying and controlling the sources of high amplitude transient noises in the OR.

The spectral content of the noise typically present in the OR during surgery, shown in Fig. 4, is relatively flat over a broad frequency range including the speech band. This fact, combined with the moderately high sustained levels, raises some concern regarding clear speech communication. In general, clear speech communication requires at least a 15 dB signal to noise ratio. Given the sustained sound pressure levels in the OR during surgery, that would suggest speech levels at 70–85 dB(A), which is well elevated from normal speaking levels of roughly 55–65 dB(A). Thus, medical staff are confronted with the choice of either speaking loudly in the OR in order to ensure good communication, or running the risk of somewhat compromised communication clarity. Given the nature of surgery, the latter of these options is clearly unacceptable. Further, since this analysis is based on L_{eq} values, it does not consider the impact of peaks in the noise rendering speech temporarily more difficult to understand. Thus, we believe that work to reduce the sustained levels of noise in ORs is important. We recommend further that any future investigation of speech communication in the OR ought to use a host of additional acoustic measures, such as the L_{50} or the speech transmission index, and focus on quicker detection times (fast or impulse settings on the sound level meter) as they are more appropriate for speech.

A limitation of this study was the absence of information linking specific acoustical events, such as use of a bone saw, to the observed high sound pressure levels. In the future, we intend to conduct measurements in the OR which either are manned by personnel noting events, or use an artificial head to record the sound simultaneously for later identification of events.

The experimental data presented here are limited to a single hospital and cannot be simply extrapolated to all hospitals. However, a previous study of hospital noise shows a rather tight distribution of sound pressure level regardless of type of hospital or geographic location.[3] Thus, it is possible that the results obtained here for operating rooms apply fairly broadly. Work is needed to confirm or deny this hypothesis.

[1]C. Baker, "Sensory overload and noise in the ICU: Sources of environmental stress," Crit. Care Nurs. Q. **6**, 66–79 (1984).

[2]A. Turner, C. King, and J. Craddock, "Measuring and reducing noise," Hospitals **49**, 85–90 (1975).

[3]I. Busch-Vishniac, J. West, C. Barnhill, R. Hunter, D. Orellana, and R. Chivukula, "Noise levels in Johns Hopkins Hospital," J. Acoust. Soc. Am. **118**, 3629–3645 (2005).

[4]A. Dellon, "To rid ORs of noise pollution," RN OR **6** (1975).

[5]A. Schneider and J. Beibuyck, "Music in the operating room," Lancet **335**, 1407 (1990).

[6]R. Dodenhoff, "Noise in the orthopaedic operating theatre," Ann. R. Coll. Surg. Engl. **77**, 8–9 (1995).

[7]G. Holmes, K. Goodman, D. Hang, and V. McCorvey, "Noise levels of orthopedic instruments and their potential health risks," Orthopedics **19**, 35–37 (1996).

[8]H. Mullett, K. Synnott, and W. Quinlan, "Occupational noise levels in orthopaedic surgery," Ir J. Med. Sci. **168**, 106 (1999).

[9]K. Moorthy, Y. Munz, S. Undre, and A. Darzi, "Objective evaluation of the effect of noise on the performance of a complex laparoscopic task," Surgery (St. Louis) **136**, 25–30 (2004).

[10]K. Willett, "Noise-induced hearing loss in orthopaedic staff," J. Bone Jt. Surg., Am. Vol., **OL-73B**, 113–115 (1991).

[11]M. Bovenzi and A. Collareta, "Noise levels in a hospital," Ind. Health **22**, 75–82 (1984).

[12]N. Shankar, K. Malhotra. S. Ahuja. and O. Tandon, "Noise pollution: A study of noise levels in the operation theatres of a general hospital during

various surgical procedures," J. Indian Med. Assoc. **99**, 244–247 (2001).

[13]B. Hodge and J. Thompson, "Noise pollution in the operating theatre," Lancet **335**, 891–894 (1990).

[14]H. Love, "Noise exposure in the orthopaedic operating theatre: A significant health hazard," Aust. N. Z. J. Surg. **73** 836–838 (2003).

[15]V. Murthy, S. Malhotra, I. Bala, and M. Raghunathan, "Detrimental effects of noise on anesthetists," Can. J. Anaesth. **42**, 608–611 (1995).

[16]M. Nott and P. West, "Orthopaedic theatre noise: A potential hazard to patients," Anaesthesia **58**, 784–787 (2003).

[17]M. Wallace, M. Ashman, and M. Matjasko, "Hearing acuity of anesthesiologists and alarm detection," Anesthesiology **81**, 13–28 (1994).

[18]B. Berglund, T. Lindvall, and D. S. (ed.), "Guidelines for community noise," Technical Report, World Health Organization (1995).

Noise in contemporary neonatal intensive care[a)]

Amber L. Williams[b)]
University of Texas Medical School-Houston, Center for Clinical Research and Evidence-Based Medicine, MSB 2.106, 6431 Fannin Street, Houston, Texas 77030

Wim van Drongelen
University of Chicago, Department of Pediatrics, Computation Institute, Chicago, Illinois 60637

Robert E. Lasky
University of Texas Medical School-Houston, Center for Clinical Research and Evidence-Based Medicine, MSB 2.106, 6431 Fannin Street, Houston, Texas 77030

Weekly sound surveys ($n=63$) were collected, using 5 s sampling intervals, for two modern neonatal intensive care units (NICUs). Median weekly equivalent sound pressure levels (L_{EQ}) for NICU A ranged from 61 to 63 dB (A weighted), depending on the level of care. NICU B L_{EQ} measurements ranged from 55 to 60 dB (A weighted). NICU B was recently built with a focus on sound abatement, explaining much of the difference between the two NICUs. Sound levels exceeded 45 dB (A weighted), recommended by the American Academy of Pediatrics, more than 70% of the time for all levels of care. Hourly L_{EQ}s below 50 dB (A weighted) and hourly L_{10}s below 55 dB (A weighted), recommended by the Sound Study Group (SSG) of the National Resource Center, were also exceeded in more than 70% of recorded samples. A third SSG recommendation, that the 1 s L_{MAX}, should not exceed 70 dB (A weighted), was exceeded relatively infrequently ($<$11% of the time). Peak impulse measurements exceeded 90 dB for 6.3% of 5 s samples recorded from NICU A and 2.8% of NICU B samples. Twenty-four h periodicities in sound levels as a function of regular staff activities were apparent, but short term variability was considerable. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717500]

I. INTRODUCTION

Concern with excessive noise levels in neonatal intensive care units (NICUs) has grown considerably over the last 30 years. Elevated noise levels may have harmful physiological consequences such as sleep disturbance, stress responses, and hearing loss.[1] Very premature infants requiring treatment in NICUs are at risk for poor developmental outcomes including retarded growth, cognitive and attention disorders, and speech and language problems.[2] Significant bilateral hearing loss affects 2–4% of NICU graduates, ten times more prevalent than healthy newborns.[3] Noise levels in NICU environments may contribute to these detrimental outcomes for high-risk infants.

In 1974, the Environmental Protection Agency (EPA) recommended sound levels of no more than 45 dB (A weighted) during the day and 35 dB (A weighted) at night to reduce the risk of noise exposure to all hospital patients.[4] The American Academy of Pediatrics (AAP) supports the recommendation from the EPA, emphasizing that noise levels above 45 dB (A weighted) should be avoided in NICUs.[5] The Sound Study Group (SSG) of the National Resource Center recommends 1) hourly equivalent sound levels (L_{EQ}) below 50 dB (A weighted) in NICUs, 2) hourly L_{10}s below 55 dB (A weighted), and 3) a 1 s L_{MAX} of no more than 70 dB (A weighted).[6] The World Health Organization (WHO) recommends that daytime noise levels in patient treatment rooms should not exceed 35 dB (A weighted).[7]

The poor outcomes of NICU graduates have stimulated a patient-based approach to NICU design and management, including greater attention to noise control.[8] Many hard surfaces have been replaced by sound absorbent material, new incubators produce less noise, and highly traveled areas are separated from patient beds.[9] However, other changes in neonatal care have increased the noise levels in contemporary NICUs, including the widespread use of cellular telephones, alarm systems for monitoring patients, and high-frequency ventilation.[10] Despite the concern for noise in NICUs, daytime and nighttime noise levels in hospitals have been steadily rising over the past 40 years.[11]

Most published NICU sound surveys do not describe modern, newly designed units. They are based on a small number of recordings and cannot detect patterns over time.[12–17] The modern NICU is a complex heath care facility with diverse sound environments that vary by the extent of patient care. Informative surveys must include sufficient sampling to adequately characterize that diversity. Previous authors have also noted that many published surveys in the medical literature inaccurately calculate overall sound levels.[11,18] By taking the arithmetic average of sound mea-

TABLE I. Mean (SD) room characteristics for each survey type.

Room	Patients per Room	IV Pumps	Conventional Ventilators	High-Frequency Ventilators
Isolation A	1.11[a]	1.22	0.78	0.11
(*n*=9)	(0.33)	(1.09)	(0.67)	(0.33)
Isolation B	1	0.40	0.50	0.20
(*n*=10)	(0)	(0.52)	(0.53)	(0.42)
Level II A	5.20	0.90	0	0
(*n*=10)	(1.03)	(1.2)	(0)	(0)
Level II B	7.20	0.30	0	0
(*n*=10)	(1.03)	(0.48)	(0)	(0)
Level III A	6.90	1.40	2.40	0.20
(*n*=10)	(1.91)	(1.71)	(1.43)	(0.42)
Level III B	6.50	2.07	2.00	0.14
(*n*=14)	(1.16)	(1.54)	(1.47)	(0.36)

[a]Two patients shared a room during one survey in Isolation A.

surements in decibels rather than converting from the logarithmic dB scale to sound pressures in Pascals, previous studies have underestimated the noise exposure for newborns in NICUs.[12–17]

This study compares noise levels in two NICUs from two different hospitals. The sound surveys recorded were further classified according to the level of care provided: Isolation, Level II, or Level III. Patients in Level II rooms usually do not require mechanical ventilation or extensive medications. Level III patients are critically ill and require more intensive care than patients in Level II. Isolation rooms are reserved for patients with infectious diseases or conditions requiring separation from other patients. The surveys in this study were recorded for seven days to ensure that diurnal variations in sound levels could be assessed. In addition, multiple surveys were collected to adequately evaluate sampling variation and generalizability of the results.

II. METHODOLOGY

A. NICU environment

NICU A was a 56 bed unit in a large county hospital in Harris County, Houston, Texas. NICU B was a 104 bed unit in a large private hospital also in Harris County, Houston, Texas. Both hospitals were teaching hospitals associated with the University of Texas Houston Medical School. The physicians (attending neonatologists, specialists, fellows, and residents) were from the University of Texas Houston Medical School. They rotated through both hospitals. Medical care, including rounding, was similar at the two hospitals. The nurses, respiratory therapists, ward clerks, and medical technicians were employees of their respective hospitals. The ratio of nurses and babies were similar in both NICUs, following AAP *Guidelines for Perinatal Care*.[19] Three to four newborns in Level II care were assigned to each nurse, whereas each nurse in Level III cared for one or two infants.

NICU A had one large room for Level III care (14 beds), six Level II rooms (six beds each), and six isolation rooms. It was constructed in 1989. The floors were vinyl throughout the unit; the walls were painted sheetrock with metal studs, and the ceilings were acoustic tiling. NICU B was opened in August of 1999. Noise control was a priority in designing

NICU B. Level II and Level III rooms had carpeted floors and contained eight beds in each room. Level II rooms were larger than rooms in Level III. Patient rooms were separated from staff workstations, storage for linens and supplies, sinks, and connecting hallways. The Isolation rooms had commercial tile floors because hard surfaces were required for easy disinfecting. Incubators, ventilators, and other biomedical equipment were similar in both NICUs. Summary statistics describing the typical room characteristics for each level of care are presented in Table I. Schematics of the bed spaces and workstations in the three room types are presented for both NICUs in Fig. 1. (Drawings are not to scale.)

B. Sample

The daily population in NICU B was approximately three times larger than the daily population in NICU A. In 2003, NICU A admitted 246 patients, and NICU B admitted 935 patients. The mean length of stay was approximately the same for both hospitals, NICU A averaged 31 days, and NICU B averaged 35 days. A summary of population characteristics for both NICUs is presented in Table II.

In NICU A, 29 surveys were recorded: nine Isolation, ten Level II, and ten Level III. The recordings in NICU A were collected between September 2005 and December 2005. In NICU B, 34 surveys were recorded: ten Isolation, ten Level II, and 14 Level III. The sound surveys in NICU B were recorded from December 2004 to May 2005.

C. Instrumentation

The sound surveys were recorded by Larson Davis Spark® 703+ Personal Noise Dosimeters, and evaluated with Larson Davis Blaze® Software (Larson Davis, Provo, Utah). A Larson Davis System 824 sound level meter (which was calibrated using a Brüel & Kjær pistonphone) was used to calibrate the 703+ dosimeters. L_{EQ} measurements on the sound level meter were within 1 dB (A weighted) of the measurements on the dosimeters. A Larson Davis Spark® 706 RC dosimeter was used to program the 703+ dosimeters prior to placement in the NICU. AC adaptors were added to the 703+ dosimeters to prevent insufficient battery power from interrupting a recording.

NICU A

FIG. 1. (Color online) Schematic drawings of the floor plans for the two NICUs (not drawn to scale).

TABLE II. Patient population characteristics for NICU A and NICU B, based on data from 2003.

	NICU A	NICU B
Number of patients	246	935
Less than 1500 g birthweight	37%	30%
Less than 1000 g birthweight	20%	17%
Mean daily census	29	88
Mean length of stay	31	35
Requiring mechanical ventilation	67%	50%
Racial/ethnic characteristics:		
Hispanic	90%	24%
African-American	7%	47%
White	2%	23%
Asian	1%	4%
Other/Unknown	0%	2%

recordings were made with a slow exponential time constant (1 s) in order to compare the measured sound levels to recommended levels from the literature.[6]

D. Procedure

Prior to recording, a map of relevant sound sources was drawn for each patient care room. The maps illustrated the placement of the dosimeter within the room as well as the locations of ventilators, pumps, humidifiers, music sources, sinks, trash and linen containers, drawers, and Extra Corporeal Membrane Oxygenation patients. The number of patients and staff present in the room were also recorded. In NICU B the microphones were suspended from unused cabinets with the body of the dosimeter in the cabinet. For recordings in NICU A, microphone stands were designed to allow microphone placement comparable in height and distance from the babies as the placement in NICU B. The microphones were positioned slightly above and behind the patients' heads, approximately 4 ft. away (about eye level for an adult). The dosimeter was retrieved eight days after the recording began.

E. Analyses

Because the setup day and retrieval day of the surveys were not a complete 24 h, they were not included in the analyses. The remaining record for each survey was exactly seven days long, beginning at midnight on the first day and ending at midnight on the seventh day, for $n = 120,960$ 5 s samples per survey.

For each survey, L_{EQ} and L_N measurements were back transformed to sound pressure levels in Pascals in order to ensure interpretable group statistics comparing the distributions of sound levels by NICU, room type, and over time. Sound level differences in Pascals in both NICUs, the three room types, and across sound level percentiles (L_Ns) were analyzed by randomization tests on the F statistic from factorial analyses of variance (ANOVAs). Heteroscedasticity was a concern in these analyses that could not be entirely

For each recording, the dosimeters calculated four sound level measurements at intervals of 5 seconds. The rms measurements, 1) equivalent sound level (L_{EQ}), 2) maximum sound level (L_{MAX}), and 3) minimum sound level (L_{MIN}), for each interval were measured in A-weighted dB (hereafter abbreviated dB(A)), and the peak sound level (L_{PEAK}) was measured in dB sound pressure level (SPL) (unweighted). The dosimeters also calculated A-weighted percentile values over longer intervals of 5 min. These values, represented by L_N, are equivalent to the sound level that was exceeded $N\%$ of the time during that interval. L_{10}, L_{30}, L_{50}, L_{70}, and L_{90} were recorded at 5 min intervals during the surveys. The dosimeters' gain was set to the maximum level, 30 dB, producing the lowest possible measurement ranges. The rms measurement range extended from 43 to 113 dBA (re 20 μPa), and the peak detector measurement range extended from 80 to 116 dB SPL. The instrument noise floor was 35 dB SPL. These ranges were determined to be appropriate to the sound levels in the NICUs during pre-testing. The

addressed by transforming the data nor by using nonparametric statistics such as Kruskal-Wallis ANOVAs that do not assume normally distributed data but do assume similarly distributed data across the groups compared. Therefore, randomization ANOVAs (10,000 Monte Carlo samples) were calculated. Randomization tests make no assumptions about the distributions compared.

Due to familiarity with the dB scale, summary statistics are presented in the dB scale (re 20 μPa) by transforming the sound pressures analyzed in Pascals. We will present medians as measures of central tendency and the 90th–10th percentile intervals as measures of dispersion because of the skewed distributions and because percentiles are not altered by log transformations. The mean and standard deviation (SD) of sound pressure values are not the same as the mean and SD of sound levels in dB. The median and percentile range, however, correspond to the same sound levels whether calculated in Pascals or in dB.

Autocorrelation functions were also calculated on the back transformed sound pressures in Pascals. Sound levels in dB may more directly reflect human perceptual discrimination, possibly providing information that is difficult to evaluate when sound measurements in Pascals are analyzed. Therefore, we repeated the autocorrelation analyses on the sound measurements in dB. The results were similar.

Autocorrelation coefficients were the dependent variables in one sample t tests to determine whether 24 and 12 h periodicities were consistently present ($r \neq 0$) in each NICU/room-type combination. The correlation coefficients were also the dependent variables in randomization ANOVAs (two NICUs $\times$ three rooms) to determine whether those periodicities varied by NICU and room type. Randomization ANOVAs were also calculated using autocorrelation time constants to determine if there were significant NICU/room differences.

The analyses were conducted using STATA statistical software package (version 9.0; May, 2005; StataCorp., College Station, TX, 77845). Amplitude spectra and autocorrelation of the demeaned time series (i.e., autocovariance function) were computed using MATLAB (version 7.0.1; The Mathworks, Inc. Natick, MA, 01760). The time constant characterizing the main peak in the autocorrelation (around zero delay) was determined as the delay at a 0.37 correlation.

III. RESULTS

A. Noise levels

Representative surveys were chosen from each of the three room types in both NICUs to illustrate the distributions of 5 s L_{EQ} measurements over one week (see Fig. 2). Measured L_{EQ}s from NICU A were elevated relative to those in NICU B for all room types. Nevertheless, the differences among the three room types were similar across the two NICUs. As expected, all the distributions were positively skewed; this was readily apparent even when plotted on a logarithmic (dB) scale. The dispersion of the distributions differed by room type. There was an obvious peak in the L_{EQ} values from the Isolation rooms that was less extreme in Level III and difficult to identify in Level II. Over 25% of
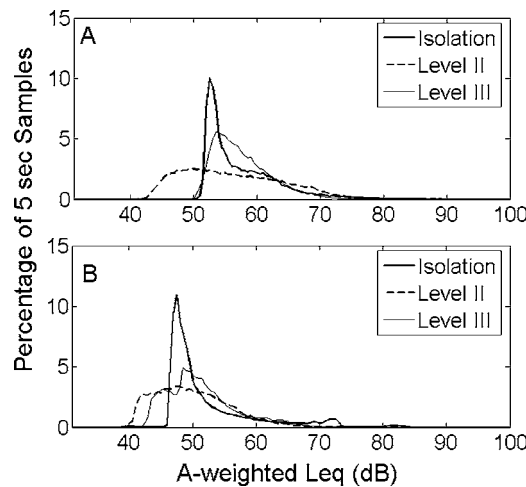
FIG. 2. The distribution of A-weighted L_{EQ} values from representative surveys in NICU A and NICU B.

the measured L_{EQ} values in both NICU Isolation rooms were within 0.5 dB of the mode for that room. The modal peaks in the Isolation rooms were also within a few dB of the quietest intervals recorded in those rooms. Thus, most of the time in the Isolation units the noise levels were near the quietest levels recorded in those rooms: approximately 50 dBA in NICU A and 45 dBA in NICU B. The Level II distributions were the most dispersed (least kurtotic), at times very noisy, but often the quietest of the three room types.

Although representative, the distributions in Fig. 2 are single surveys. Data from all the surveys are presented in Fig. 3 by calculating overall L_{EQ} values for each seven day survey. As expected, the overall L_{EQ}s from NICU A were consistently higher than those recorded from the corresponding levels of care in NICU B ($p < 0.001$). However, there were striking differences in seven day L_{EQ} values in the three rooms depending on the NICU (test for interaction: $p < 0.001$). In NICU B, Level II rooms were less noisy than the other two room types. In contrast, Level II rooms were the noisiest rooms in NICU A. Further insight into these results is provided by Fig. 4, presenting the seven day L_Ns. Isolation rooms and Level III rooms from each NICU had similar noise levels at all noise percentile levels except L_{10}, which were higher in the Isolation rooms. Relative to Isolation and Level III L_Ns, Level II L_Ns were dramatically different in NICUs A and B. In NICU B, Level II rooms were
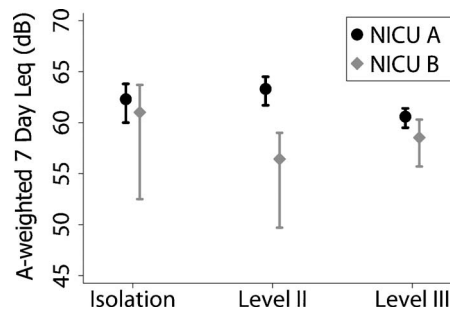
FIG. 3. Median A-weighted 7 day L_{EQ} measurements for each NICU and room type (error bars represent the range).
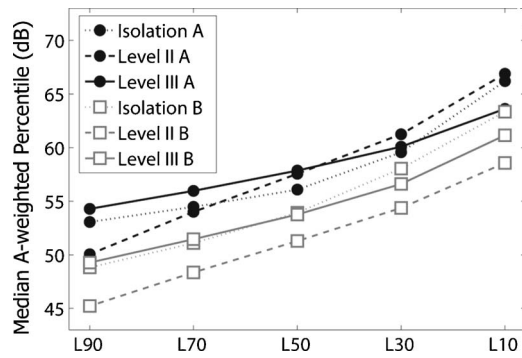
FIG. 4. Median A-weighted 7 day L_N values for each NICU and room type.

the quietest at all noise percentiles. In NICU A, Level II rooms were the quietest for L_{90} but the loudest for L_{10}.

Variability among surveys in overall L_{EQ} measurements was greater in NICU B than NICU A, especially in Isolation and Level II. The highest variability among surveys in overall L_{EQ} measurements for both hospitals was in the Isolation rooms, followed closely by Level II rooms (see Fig. 3). Although error bars are omitted in Fig. 4 for clarity, variability also tended to be greatest in Isolation units followed by Level II rooms at all percentiles (L_Ns) recorded.

B. Peak sound levels

Because loud sounds are most likely to be associated with adverse outcomes, the loudest sounds were further evaluated by analyzing peak sound levels. For each 5 s sampling period, the loudest peak sound (regardless of its duration) was recorded by the dosimeters. Peak sound levels were divided into three categories: 1) below 80 dB SPL, 2) 80–90 dB SPL, and 3) above 90 dB SPL. These three categories were chosen because 80 dB represents the lowest accurate peak measurements on the dosimeters, and L_{PEAK} values above 90 dB are intense sounds that may be especially disruptive. Figure 5 illustrates the percentage of peak measurements from the three peak level categories for each

of the three room types in both hospitals. Percentages were calculated from 120,960 5 s samples during each week of recording.

There were no significant differences between the NICUs or room types for the two lowest peak categories, <80 dB and 80–90 dB. However, there were striking room differences that varied by NICU in the highest peak category, >90 dB (test for the NICU X room-type interaction: $p = 0.001$). The highest peak sound levels (>90 dB) were recorded more often in all three room types from NICU A ($p < 0.001$). While the differences were not statistically significant, the lowest peak sound levels (<80 dB SPL) were more frequent in NICU B Level II and Level III rooms compared to NICU A Level II and Level III rooms. In contrast, the reverse characterized the Isolation rooms in the two NICUs.

C. Adherence to recommendations

The sound samples were analyzed to evaluate adherence to the American Academy of Pediatrics (AAP) and Sound Study Group (SSG) recommendations. We did not directly evaluate adherence to the WHO recommendation that sound levels not exceed 35 dBA because that sound level was below the limit of detection for the operative range of the dosimeters used in this study. Indirectly, the WHO recommendation was exceeded more often than the more liberal AAP and SSG recommendations.

The recommendation from the AAP is that sound levels should not exceed 45 dBA. The SSG recommendations are 1) hourly L_{EQ} should not exceed 50 dBA, 2) hourly L_{10} should not exceed 55 dBA, and 3) 1 s L_{MAX} should not exceed 70 dBA. Figure 6 presents the percentage of samples that met the recommendations from the AAP and the SSG.

Rarely were the recommendations met, with the exception of the third SSG recommendation. Recommended sound levels were achieved more frequently in NICU B than in NICU A (all p values<0.001). Excluding SSG 3, the other three recommendations were more often met in Isolation and Level II than in Level III (AAP $p < 0.001$; SSG1 $p = 0.002$; SSG2 $p = 0.004$).
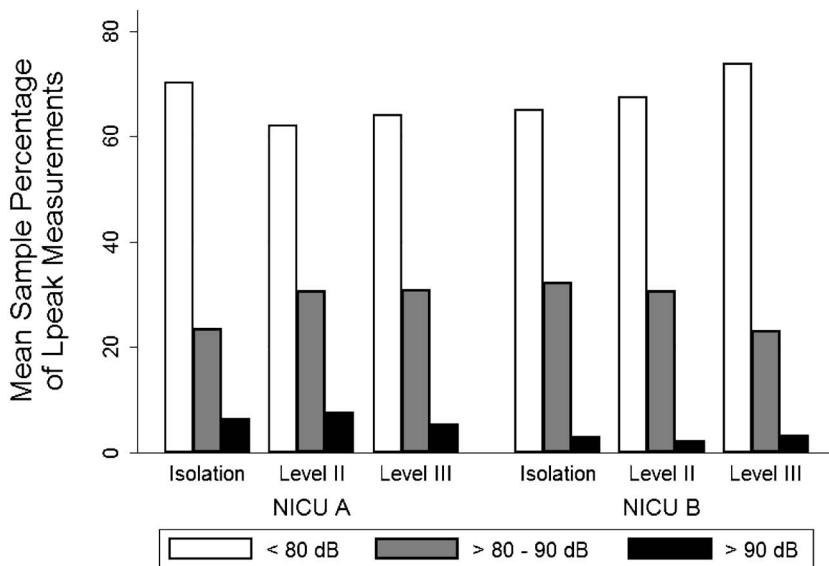
FIG. 5. Mean percentage of L_{PEAK} measurements in three categories: <80 dB, 80–90 dB, and >90 dB.
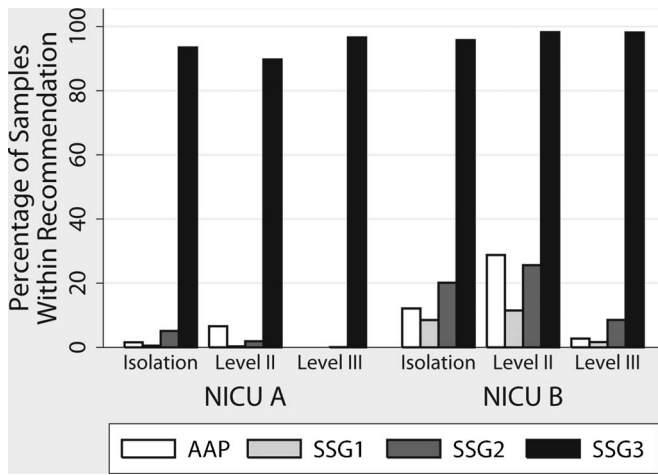
FIG. 6. Mean percentage of sound level samples within the AAP recommendation: A-weighted sound levels should not exceed 45 dB, and the three Sound Study Group recommendations: 1) hourly A-weighted L_{EQ} should not exceed 50 dB, 2) hourly A-weighted L_{10} should not exceed 55 dB, and 3) 1-second A-weighted L_{MAX} should not exceed 70 dB.

FIG. 7. Weekly time series of median A-weighted L_{EQ} values in 2 hour intervals for the three room types in both NICUs.

Three of the recommendations presented in Fig. 6 were calculated directly from the data recorded in the NICUs. Adherence to SSG 3 could not be directly calculated because our minimal sampling intervals were 5 s, and the recommendation applies to 1 s sampling. However, we estimated from our data a minimum percentage of time that L_{MAX} was below 70 dBA by calculating the percentage of 5 s L_{MAX} measurements that did not exceed 70 dBA. Because L_{MAX} measures the maximum sound level in an interval, we can assume that if the 5 s L_{MAX} does not exceed 70 dBA, then none of the 1 s L_{MAX} measurements in the same interval exceeded 70 dBA.

Contrasting with the results for the other recommendations, sound levels rarely exceeded 70 dBA in any room in either NICU. That may be surprising given Fig. 5 that indicates peak sound levels always exceeded 70 dB SPL. The difference is that the L_{MAX} measurement in SSG 3 integrates sound levels over 1 s, and the peak sounds were measured using a separate circuit with a 50 μs rise time.

D. Variability over time

The remaining analyses focus on the variations in sound levels over time. Variability in sound levels may better predict arousal and behavioral state changes in newborns than the measured sound levels reviewed above. Sounds that are constant over time are perceived as less disrupting than random and highly variable sounds.[20–22]

Weekly and daily periodicities were analyzed to determine whether high and low noise levels reoccurred at the same times. For the following graphical representations, "median weeks" were constructed for each room type in each NICU. For each 5 s interval throughout the week, the median L_{EQ} was chosen from the group of surveys of the same NICU and room type. Figure 7 presents the time series of the median L_{EQ}s for each NICU and room type over 2 h intervals. Diurnal variations in sound levels are evident in all three rooms and both NICUs. As expected, sound levels were highest during the workday (between 7:00 a.m. and 7:00
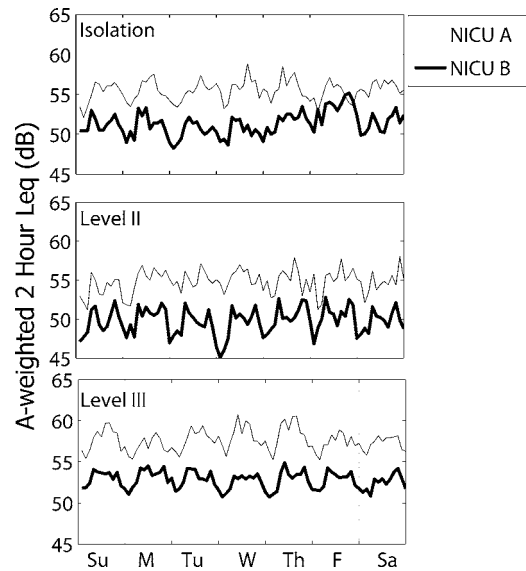
p.m., as indicated below) and dropped at night and in the early morning by up to 5 dB. Differences from day to day were less obvious.

A "median day" was constructed in a similar manner for all 5 s L_{EQ} values. All surveys for the same NICU and room type were further separated into seven daily surveys beginning at 12:00 a.m., and the median L_{EQ} for each 5 s interval was chosen from all daily surveys for each room type in both NICUs. Figure 8 presents median daily time series for 5 min intervals in order to detail the periodicities throughout each day. The data are presented (but not analyzed) for 5 min intervals to reduce the random variability of the data at shorter time intervals that obscured the real periodic trends in the data. There was a daily peak in sound levels in both NICUs around 7:00 a.m. (slightly later in NICU A). In NICU B there was another peak at around 7:00 p.m. There were additional peaks (more obvious in NICU A) at 9:00 a.m., 12:00 p.m., and 3:00 p.m. NICUs were quietest after midnight through the early morning hours. The two prominent
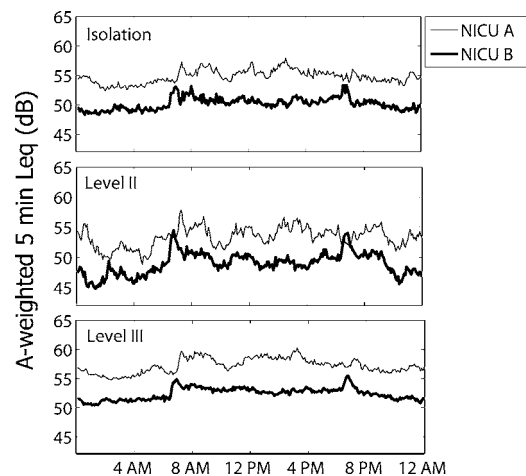
FIG. 8. Daily time series of median A-weighted L_{EQ} values in 5 minute intervals for the three room types in both NICUs.

Williams *et al.*: Neonatal intensive care noise
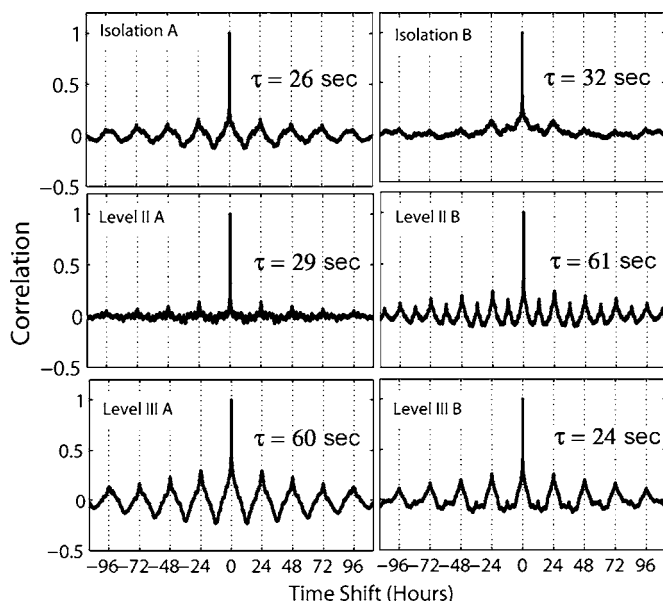
FIG. 9. Autocorrelation functions from 5 second A-weighted L_{EQ} "median week" surveys.

FIG. 10. Autocorrelation functions from 5 minute A-weighed L_{90}-L_{10} range "median week" surveys.

peaks, at 7:00 a.m. and 7:00 p.m., corresponded to the nursing staff shift changes. At those times each baby was assessed, cared for, and signed off to the nurse responsible for the patient during the next shift. Shift changes are recognized as busy and noisy times in NICUs; 9:00 a.m., 12:00 p.m., and 3:00 p.m. were times nurses recorded vital signs, changed diapers, initiated feedings, and otherwise cared for their patients.

Percentile range values were chosen to describe patterns of variation in sound levels over time. A "median week" was also generated for the ranges between the 90th and 10th percentile measurements. For each 5 min interval, the median range (L_{10}–L_{90}) was calculated from each group of surveys for the same hospital and room type.

We calculated autocorrelation (autocovariance) functions and amplitude spectra on the time series of the median week L_{EQ} values (Fig. 9) and the median L_{90}–L_{10} ranges (Fig. 10) for the three room types in each NICU. Spectrum and autocorrelation are equivalent in the frequency and time domain, respectively. We present the autocorrelation functions because they more directly communicate the relationship between the samples in our data.

The autocorrelation functions of the median L_{EQ} time series (Fig. 9) show that a 24 h periodic component was present in all traces from both NICUs. The strength of the 24 h component varied among room types and between the NICUs. In order to quantitatively characterize the periodicity at 24 h, we calculated autocorrelation functions for the individual surveys. From the 63 individual surveys we measured the correlation coefficient (r) at 24 and 12 h. The randomization ANOVA indicated a significant interaction between room type and NICU for the 24 h L_{EQ} correlation coefficient ($p<0.001$). The 24 h correlations were similar in Isolation rooms in both NICUs. In Level II, NICU B had stronger correlations at 24 h than NICU A. For Level III rooms, however, NICU A had higher L_{EQ} correlations at 24 h compared to NICU B.
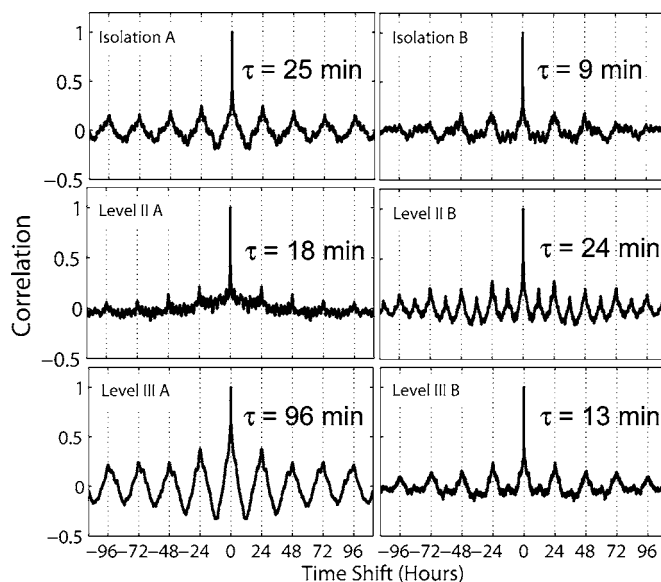
T tests confirmed that 24 h L_{EQ} correlation coefficients in all room types from NICU A were unlikely to be zero (Isolation $p=0.010$; Level II $p=0.008$; Level III $p<0.001$). L_{EQ} correlation coefficients from NICU B were unlikely to be zero in Level II ($p<0.001$) and Level III ($p=0.002$) and just missed significance at the 0.05 level for Isolation ($p=0.060$). Over a maximum lag of 96 h, the mean correlation coefficients at 24 h intervals ranged from 0.013 in NICU A Level II to 0.050 in NICU A Level III.

L_{EQ} correlation coefficients at 12 h were modest at best. The t tests for the 12 h L_{EQ} correlation coefficients indicated that only two groups had significant nonzero values, and in one of those room types, the correlation was negative: NICU A Level III (mean $r=-0.02$, $p=0.030$) and NICU B Level II (mean $r=0.03$, $p<0.001$). The L_{EQ} correlation coefficients at 12 h were consistently higher in NICU B than in NICU A ($p<0.001$). In both NICUs, 12 h L_{EQ} correlations were highest in Level II compared to Level III and Isolation (NICU A: $p=0.006$; NICU B $p=0.030$).

The L_{EQ} time constant (τ) associated with the exponential decay around 0 delay varied between 9 and 219 s in the full sample of surveys, depending on the room and NICU. However, these room/NICU differences in time constants were not statistically significant due to the variability among surveys. The relatively steep decline in the autocorrelation functions at small delays indicates there is a strong random component in the time series. The exponential decay can be interpreted as the component of a Markov process or as a filter applied to the underlying random process. We may conclude that in addition to a small but consistent periodicity every 24 h, a random sound generation process dominates the measurements.

The autocorrelation functions for the median week of L_{90}–L_{10} ranges are shown in Fig. 10. Autocorrelations of the L_{90}–L_{10} range median week surveys also shows 24 h periodicity in combination with a pronounced exponential decay at 0 delay (Fig. 10). Mean correlation coefficients at 24 h

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Williams *et al.*: Neonatal intensive care noise   2687

ranged from 0.02 in NICU B Isolation to 0.12 in NICU A Level III. One sample t tests confirmed that 24 h periodicities for $L_{90}-L_{10}$ ranges are present for all levels of care in NICU A (Isolation $p=0.009$; Level II $p<0.001$; Level III $p<0.001$), and in NICU B Levels II and III (Level II $p=0.004$; Level III $p=0.004$). Twenty four hour correlations for $L_{90}-L_{10}$ ranges in NICU B Isolation were not consistently present ($p=0.130$). Only NICU A Level III showed a significantly larger $L_{90}-L_{10}$ range correlation at 24 h than other rooms ($p<0.001$).

$L_{90}-L_{10}$ range correlations at 12 h ranged from -0.03 in NICU A Level III to 0.06 in NICU B Level II. The 12 h correlations were consistently nonzero in three room types: NICU A Level II (mean $r=0.03$, $p=0.049$), NICU B Isolation (mean $r=0.03$, $p=0.050$), and NICU B Level II (mean $r=0.06$, $p=0.020$).

Time constants (τ) for $L_{90}-L_{10}$ range autocorrelation functions ranged from 4 min to 178 min in the full sample of surveys. The longer time constants than reported for the L_{EQ} autocorrelation functions were expected since the ranges for $L_{90}-L_{10}$ were calculated over 5 min intervals, rather than 5 s. Room/NICU differences in $L_{90}-L_{10}$ range time constants were not statistically significant due to the variability among surveys.

IV. CONCLUSIONS

A. Adherence to recommendations

The two NICUs surveyed are representative of large, busy, state-of-the-art units. One NICU was recently constructed and designed with a focus on noise abatement. The other was built when noise control was a lesser priority, however the unit has been recently updated. Most of the noise levels sampled from both NICUs exceeded recommendations from the two most widely recognized advisory groups in the U.S. and by inference, the stricter WHO recommendation (see Fig. 6). The one exception was that the 1 s L_{MAX} was rarely greater than the recommended 70 dBA.[6] Modern large, busy NICUs are noisy workplaces and rarely are as quiet as expert panels recommend for optimal patient care. On the other hand, sustained sound levels above 70 dBA are also rarely encountered, although transients frequently exceed 70 dB SPL (see Fig. 5). Because the AAP recommended sound level of 45 dBA approaches the lower limit of measurement on the dosimeters (40 dBA), our results may slightly overestimate the amount of time the NICUs exceeded this level, especially in the quietest rooms (NICU B Level II). WHO recommends daytime hospital noise levels of 35 dBA and nighttime levels of 30 dBA, both dramatically lower than recommendations by the AAP and SSG. We did not directly calculate adherence to WHO guidelines due to the lower measurement limit on the dosimeters, but considering the low adherence to AAP and SSG guidelines, sound levels meeting WHO recommendations would be less frequent.

B. NICU comparison

There were consistent differences in sound levels between the two NICUs. Not unexpectedly, noise levels in the recently constructed NICU B were lower than NICU A, especially in Level II rooms. On the other hand, the variability in sound levels among surveys was greater in NICU B than NICU A. The most obvious explanation for the difference is the attention to noise abatement in the construction of NICU B. In particular, separating spaces for activities that do not require interacting directly with the patients from the patient locations, the use of carpeting rather than industrial tile on the floor, and the use of other sound attenuating materials (e.g., plastic drawers and trash cans) are likely to be responsible. Parent and staff activities may also have contributed to the NICU differences.

C. Level of care

The differences in noise levels among NICUs cannot be understood without also considering the level of care and, therefore, the room type. Quiet intervals in isolation units presumably reflect intensive care equipment noise (ventilator assistance, feeding pumps, neonatal monitors, etc.) when the patients are alone in their rooms. The relatively infrequent but much louder noise levels were presumably recorded when medical staff and parents entered the Isolation rooms or when alarms were triggered on biomedical equipment.

Level III care is often similar to that in Isolation, in terms of noise generating instrumentation, but differs because other patients and medical staff are always present. That difference could explain much of the disparity between the Isolation and Level III distributions.

As expected, the quietest recordings were from Level II rooms in both NICUs (Figs. 2, 4), probably because of the reduced need for intensive care. However, at times Level II rooms were also the noisiest rooms. Thus, they tended to be the most variable over time (Figs. 2, 7, 8). Despite these similarities, there were striking differences between NICU A and NICU B Level II sound levels. Level II rooms in NICU A had the highest L_{EQ}s of the three room types (Fig. 3) because of their high noise levels at the loudest percentiles recorded, L_{10} and L_{30} (Fig. 4). In contrast, Level II rooms in NICU B had the lowest L_{EQ}s of the three room types.

An explanation for the discrepancy at the loudest noise levels in Level II rooms between the two NICUs is not obvious. When NICU A Level II rooms became loud, they became very loud, exacerbated by the reverberant surfaces in a relatively small room with up to six newborns. The loudest noise levels in NICU B's Level II rooms may not have been as loud as in NICU A because of the noise abating architectural acoustics and the larger room sizes. The range in medical care activities can be great for Level II patients. In addition, Level II patients may, at times, require minimal attention and have no visitors, while at other times they may receive many visitors interacting with them as newborns soon to be discharged.

D. Variability over time

Previous NICU sound surveys have emphasized the lack of variation in sound levels during the day and week.[12,23] In contrast, we observed striking differences, approximately 5 dB, during the day (Fig. 7). In accord with previous stud-

ies, differences in sound level from one day to the next were less apparent. The discrepancy between our results and those of other studies may be explained by the longer duration and greater number of surveys we recorded. Noise levels are highly variable in NICUs, and trends can be obscured in studies with small samples. Although our results indicate systematic variation in sound levels during the day, they also indicate the randomness of the observed variations in sound levels. We calculated autocorrelations on time series that sampled the sound levels at 5 s (L_{EQ}) and 5 minute ($L_{90}-L_{10}$ range) intervals. At that resolution, 24 h periodicities are reliably demonstrated, but the essential randomness of the noise is obvious. NICU activities (rounds, shift changes, parental visits, etc.) do not occur exactly at the same time each day. Thus, reliable long term variation in sound levels (on the order of hours) contrasts with the short term (minutes and seconds) randomness of the sound. It should be noted that the short term "random" fluctuations in sound levels can dwarf the longer term variation.

The importance of these results is how they affect the newborns in NICUs. The concern with diurnal cycles in the NICU is that sick newborns have undisturbed times to sleep, recuperate, and grow. Our data indicate there is a regular, daily cycle of noise levels in the NICU, a cycle that is imposed by staffing and visiting policies. The relevant issues for patient care are whether the diurnal variation and the more pronounced short term variations disrupt the sick newborn.

E. Strengths and limitations

This study differs from previous surveys in several important respects. One motivation for this study was that prior sound surveys of NICUs have not captured the diversity of sound environments in the contemporary unit. The modern NICU has evolved into a complex facility that can vary dramatically between units, rooms, patients, and over time. We have repeatedly sampled the sound environments in two NICUs. Both NICUs are large, busy units that are similar to other leading units in the U.S. and other countries. By conducting many week long surveys our results are unlikely to reflect the vagaries of surveying at unusual or unrepresentative times. Furthermore, because of the number of samples we recorded, we can differentiate robust effects from chance results.

This analysis also differs from other studies in reporting a number of measurements that are typically not measured or analyzed. In addition to measures like L_{EQ}, characterizing noise levels at the loudest and quietest times may prove to be very important in designing NICUs. Peak sounds may be especially problematic for sick newborns. The overall level may be less important than how much sound levels change over time. Even the sick newborn may adapt to a noisy room but may have a much harder time doing so if those noises vary significantly over time. On the other hand, newborns may tolerate well the high noise levels experienced in the NICU as long as there are periods of relative quiet for needed rest and recuperation. Periods of relative quiet may be most important for the growth and development of sick

newborns. All of these issues require further study and must be considered when making recommendations for NICUs.

Our surveys do not include observations of staff and parent behaviors or biomedical equipment during the sound recordings. We have identified events that are possibly contributing to elevated noise levels based on general observations of the different levels of care, but we cannot conclusively distinguish the sound sources responsible for these differences from our data.

Although we have captured some of the diversity in the NICU noise environment by distinguishing room type, there is further diversity at the individual patient level. For example, we did not assess noise levels in isolettes that many newborns use. Those isolettes both generate noise (temperature regulation, opening and closing of ports, etc.) and attenuate external noise. The sound exposure in NICUs cannot be understood without consideration of isolette acoustics.

Room and isolette acoustics do not fully characterize the noise exposures to NICU patients. Rather, each patient has his/her own unique exposure history depending on how sick that newborn is, the NICU resources available, the medical decisions that are made, and the family's involvement. When trying to determine the effects of NICU noise on newborns in order to identify acceptable noise levels, noise exposures of the individual newborns, and not the room, must be considered. By shifting the focus from room acoustics to infant noise exposures, other issues become important. The concern shifts to the sound each infant is exposed to and the infant's response to that sound exposure. Many of the sickest newborns in the NICU are also very premature (as young as 22–23 weeks postmenstrual age). At that age fetuses are just beginning to respond to very loud sounds,[24] and sick preterm newborns may be relatively unresponsive to sounds they encounter in the NICU. Meaningful recommendations for sound levels in NICUs depend on understanding how newborns respond to sounds and how NICU sounds influence their health and development.

F. Future directions

Noises that affect health and development in NICU newborns remain to be determined. Future sound surveys should include video recordings or observers to document specific noise events in NICUs. Ear canal acoustics and the maturation of the auditory and arousal systems must also be considered in evaluating the adequacy of the NICU sound environment. Beyond room and device acoustics, individual sound exposures must be measured in order to fully evaluate the health implications of the NICU sound environment.

In order to develop evidenced-based recommendations for NICU noise levels, information is needed regarding the NICU sound environment, the responses of sick newborns to sound, and the trade-offs between the aversive consequences of noise, and the benefits of those sounds to the patient and staff. Communication is necessary for effective medical care, noise-generating instrumentation is present for monitoring and intervention, and language and other sounds are important for the infants' social interactions and bonding. This study provides needed information characterizing the sound

levels in modern NICUs. Much additional research is needed to develop appropriate recommendations regarding sound levels in NICUs.

ACKNOWLEDGMENTS

[1] S. N. Graven, "Sound and the developing infant in the NICU: Conclusions and recommendations for care," J. Perinatol **20**, S88–S93 (2000).

[2] H. G. Taylor, N. Klein, and M. Hack, "School-age consequences of birth weight less than 750 g: A review and update," Dev. Neuropsychol. **17**, 289–321 (2000).

[3] A. Erenberg, J. Lemons, C. Sia, D. Trunkel, and P. Ziring, "Newborn and infant hearing loss: Detection and intervention," American Academy of Pediatrics, Task Force on Newborn and Infant Hearing, 1998–1999, Pediatrics **103**, 527–530 (1999).

[4] "Information on levels of environmental noise requisite to protect public health and welfare with an adequate margin of safety" (Report No. 5509-74-004), Environmental Protection Agency, Office of Noise Abatement and Control (Government Printing Office, Washington D.C., 1974).

[5] American Academy of Pediatrics Committee on Environmental Health, "Noise: A hazard for the fetus and newborn," Pediatrics **100**, 724–727 (1997).

[6] R. D. White, "Recommended standards for newborn ICU design," Committee to establish recommended standards for newborn ICU design, J. Perinatol **19**, S1–S12 (1999).

[7] "Guidelines for community noise," edited by B. Berglund, T. Lindvall, D. H. Schwela, K. T. Goh, Technical Report, World Health Organization, 1999.

[8] M. M. Shepley, "Evidence-based design for infants and staff in the Neonatal Intensive Care Unit," Clin. Perinatol. **31**, 99–311 (2004).

[9] J. B. Evans and M. K. Philbin, "The acoustic environment of hospital nurseries," J. Perinatol **20**, S105–S112 (2000).

[10] R. J. Berens and C. G. Weigle, "Noise measurements during high-frequency oscillatory and conventional mechanical ventilation," Chest **108**, 1026–1029 (1995).

[11] I. J. Busch-Vishniac, J. E. West, C. Barnhill, T. Hunter, D. Orellana, and R. Chivukula, "Noise Levels in Johns Hopkins Hospital," J. Acoust. Soc. Am. **118**, 3629–3645 (2005).

[12] A. W. Gottfried, P. Wallace-Lande, S. Sherman-Brown, J. King, C. Coen, and J. E. Hodgman, "Physical and social environment of newborn infants in special care units," Science **214**, 673–675 (1981).

[13] G. D. Levy, D. J. Woolston, and J. V. Browne, "Mean noise amounts in level II vs. level III neonatal intensive care units," Neonatal Network **22**, 33–37 (2003).

[14] W. D. T. Kent, A. K. W. Tan, M. C. Clarke, and T. Bardell, "Excessive noise levels in the neonatal ICU: Potential effects on auditory system development," J. Otolaryngol. **31**, 355–360 (2002).

[15] F. Benini, V. Magnavita, P. Lago, E. Arslan, and P. Pisan, "Evaluation of noise in the neonatal intensive care unit," Am. J. Perinatol **13**, 37–41 (1996).

[16] C. Krueger, S. Wall, L. Parker, and R. Nealis, "Elevated sound levels within a busy NICU," Neonatal Network **24**, 33–37 (2005).

[17] S. S. Surenthiran, K. Wilbraham, J. May, T. Chant, A. J. B. Emmerson, and V. E. Newton, "Noise levels within the ear and post-nasal space in neonates in intensive care," Arch. Dis. Child **88**, F315–F318 (2003).

[18] M. K. Philbin, "The full-term and premature newborn: The influence of auditory experience on behavior of preterm newborns," J. Perinatol **20**, S77–S87 (2000).

[19] American Academy of Pediatrics and American College of Obstetricians and Gynecologists, *Guidelines for Perinatal Care*, 5th ed. (American Academy of Pediatrics, Elk Grove Village, IL, 2002).

[20] J. N. Aaron, C. C. Carlisle, M. A. Carskadon, T. J. Meyer, N. S. Hill, and R. P. Millman, "Environmental noise as a cause of sleep disruption in an intermediate respiratory care unit," Sleep **19**, 707–710 (1996).

[21] N. L. Carter and H. C. Beh. "The effect of intermittent noise on vigilance performance," J. Acoust. Soc. Am. **82**, 1334–1341 (1987).

[22] J. R. Hassall and K. Zaveri, *Acoustic Noise Measurements* (Brüel and Kjær, Nærum, Denmark, 1979), Chap. 3, pp. 52–55.

[23] M. K. Philbin and L. Gray, "Changing levels of quiet in an intensive care nursery," J. Perinatol **22**, 455–460 (2002).

[24] R. E. Lasky and A. L. Williams, "The development of the auditory system from conception to term," NeoReviews **6**, 141–152 (2005).

[25] R. E. Lasky, A. L. Williams, W. van Drongelen, and L. C. Gray, "Noise in neonatal intensive care units (NICUs) and its effect on high risk newborns," J. Acoust. Soc. Am. **118**, 1956 (2005).

Chronological analysis of architectural and acoustical indices in music performance halls

Youngmin Kwon[a] and Gary W. Siebein
Architecture Technology Research Center, School of Architecture, University of Florida, 231 ARCH,
P.O. Box 115702, Gainesville, Florida 32611

This study aims to identify the changes in architectural and acoustical indices in halls for music performance built in the 18th through the 20th Centuries. Seventy-one halls are classified in five specific periods from the Classical Period (1751–1820) to the Contemporary Period (1981–2000) based on chronology in music and architectural acoustics. Architectural indices such as room shape, seating capacity, room volume, balcony configuration, and the like as well as acoustical indices such as RT, EDT, G, C80, IACC, and the like for the halls found in the literature are chronologically tabulated and statistically analyzed to identify trends and relationships in architectural and acoustical design for each of the historical periods identified. Some indices appear correlated with each other. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2713663]

I. INTRODUCTION

From Greek amphitheaters originally used for ancient theatric performances to 20th Century contemporary halls for various types of performances, architectural as well as acoustical design technologies involved in these performance venues have progressed enormously and are still evolving. Over the last century, acousticians have better understood the relations between overall acoustical qualities of the performance halls and room geometric indices such as shape, volume, dimensions, and the like. Accordingly, better room acoustical qualities have resulted from proper geometric design development.

This paper explores the changes in architectural and room acoustical indices in halls, primarily used for music performances and built in the 18th through the 20th Centuries. Seventy-one halls including fifty-nine concert halls and twelve multipurpose halls are investigated. Most were selected from *Concert Halls and Opera Houses: How They Sound* (1996) and *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (2004) by Beranek which present acoustical data of representative performance halls in the world. The halls are classified in five specific periods on the basis of chronology in music and architectural acoustics from the Classical Period (1751–1820) to the Contemporary Period (1981–2000). Architectural and acoustical indices are then chronologically surveyed and tabulated, and relationships among those indices are further investigated to identify trends in architectural and acoustical design for each of the classified historical periods. Understanding of how, over time, architectural indices of these spaces have changed in relation to room acoustical indices and how the recent design of such spaces is occurring should support both acousticians and architects in their design process.

II. RESEARCH METHOD

Seventy-one halls constructed before the 21st Century are chronologically classified in three major musical eras: Classical (1751–1820), Romantic (1821–1900), and Modern (1901–2000) eras. See Table I. The 20th Century era is subdivided into three specific periods: Pre-Modern (period of the first half a century before World War II), Modern I (period of three decades after World War II), and Modern II (contemporary period from 1981 to 2000). Talaske *et al.* (1982) and Hoffman *et al.* (2003) have remarked that "The past two decades have seen a remarkable growth in number and technical quality of facilities devoted to the performing arts" in the introductory sections for the books of *Halls for Music Performance: Two Decades of Experience, 1962-1982* (1982) and of *Halls for Music Performance: Another Two Decades of Experience 1982-2002.* (2003).

About 70% (50 halls) out of the total 71 halls were constructed within the last 50 years during the Modern I and II Periods from 1951 to 2000. However, because only one music hall is found in the Classical Period, this period is excluded from most of the following analyses.

The chronological analyses are grounded on the collected materials such as graphics, architectural drawings, and acoustical data from previously published documents and the technical literature. Specifically, architectural and acoustical indices listed in Table II are chronologically tabulated and

TABLE I. Chronological distribution of the halls studied.

Chronological period	Concert hall	Multiuse hall	Subtotal
Classical (1751–1820)	1	0	1
Romantic (1821–1900)	8	1	9
Pre-Modern (1901–1950)	9	2	11
Modern I (1951–1980)	19	4	23
Modern II (1981–2000)	22	5	27
Total	59	12	71

TABLE II. List of architectural and acoustical indices investigated.

Architectural design indices		Room acoustical indices
Room indices	Balcony overhang indices	
Room shape	Overhang depth (D)	Reverberation time (RT)
Room volume	Opening height (H)	Early decay time (EDT)
Seating capacity	Depth to height ratio (D/H)	Bass ratio (BR)
Volume per seat	Vertical angle of view (θ)	Loudness or strength (G)
Room length	Number of overhangs	Early-to-late energy index (C80)
Room width		Lateral energy fraction (LF)
Room height		Interaural cross correlation (IACC)
		Binaural quality index (BQI)
		Initial time delay gap (ITDG)
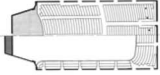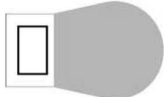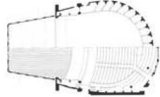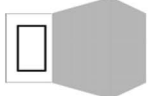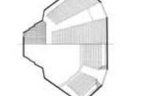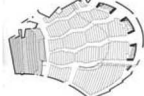
statistically analyzed to identify design trends observed in music performance hall history.

III. GEOMETRIC CLASSIFICATION METHOD

Haan and Fricke (1992) attempted to geometrically categorize auditoria in five types in their survey of auditoria: rectangular, horseshoe, fan-shaped, arena, and geometric style. The geometric classification of room shapes in this study also basically follows these five categories, but a few different terms are used: *rectangular* is replaced with *shoebox*, and *geometric* with *irregular*, because the latter terms of *shoebox* and *irregular* are commonly used in the literature. Typical configurations of the room shapes are illustrated with some examples in Table III. The room shapes are categorized primarily on the basis of architectural plan views but also

with reference to the longitudinal section views as well as the audience seating arrangements. Specifically, a hall with parallel side walls and a planar rear wall is classified in the shoebox style, a hall with oblique side walls that widen outward is classified in the fan-shaped style, and a hall with a continuously curved wall from the side to the rear is classified in the horseshoe style. A hall with a stage surrounded fully by seating areas is classified in the arena style. Note that semisurrounded halls, which can be classified as a room shape of *semiarena*, are included in the group of *arena* and statistically investigated in most analyses. Lastly, the irregular style includes a hall with a combined configuration, such as parallel side walls with a curved rear wall, and a hall with one oddly angled major surface or more. For example, Meyerson Symphony Hall in Dallas, TX, which has parallel side

TABLE III. Configurations of the room shape typologies (not to scale).

Category	Typical configuration		Example
Shoebox			Boston Symphony Hall Boston, MA, US
Horseshoe			Bass Performance Hall Fort Worth, TX, US
Fan-shaped			Patria Hall Budapest, Hungary
Arena	 Arena		Boettcher Concert Hall Denver, CO, US
	 Semi-arena		Waterfront Hall Belfast, Ireland
Irregular			Liederhalle Beethovensaal Stuttgart, Germany

* Drawings adopted from *Concert Halls and Opera Houses* by Beranek (2004).
* Halls with a stage area positioned aside from the center of the hall were subcategorized in the semiarena style.

TABLE IV. Chronological trends in room shape typologies.[b]

Room shape	Romantic (1821–1900)		Pre-Modern (1901–1950)		Modern I (1951–1980)		Modern II (1981–2000)		Overall	
	No.	%	No.	%	No.	%	No.	%	No.	%
Shoebox	6	66.7	1	9.1	8	34.8	7	25.9	23	32.4
Fan-shaped	0	0.0	6	54.5	6	26.1	4	14.8	16	22.5
Arena/Semi	1	11.1	0	0.0	5	21.7	8	29.6	14	19.7
Horseshoe	2	22.2	4	36.4	1	4.3	3	11.1	10	14.1
Irregular	0	0.0	0	0.0	3	13.0	5	18.5	8	11.3
Total	9	100.0	11	100.0	23	100.0	27	100.0	71	100.0

Note that the total number of shoebox halls (23) includes 1 shoebox hall found in the Classical Period.

walls with an opera house style curved rear wall, and Liederhalle Beethovensaal in Stuttgart, Germany, which has an inversely curved side wall, are classified in this irregular style.

IV. CHRONOLOGICAL PHASES IN ARCHITECTURAL DESIGN INDICES

In the following discussion of trends in room shape, seating capacity, volume, volume per seat, and dimensions, some statistical analyses exclude Royal Albert Hall in London or Music Shed in Lenox, MA, or both halls, because these halls are considered statistical outliers in some cases.

A. Trend in room shape typologies

As shown in Table IV, the halls are categorized in five primary room shape typologies across the chronological periods defined earlier. Shoebox halls are found to be the most widely used room shape for music performances across all time periods. A larger number of shoebox halls are observed in all periods except the Pre-Modern Period (1901–1950). Two more specific trends may further extend this discussion of trend in room shape. One is that fan-shaped halls were developed during the Pre-Modern Period and became another primary room shape. For visual as well as acoustical intimacy, fan-shaped halls were a widely used room shape for multipurpose halls from the Pre-Modern Period to the present time. In fact, all the multipurpose halls (12 halls) investigated here were built since the Pre-Modern Period and 7 halls (about 58%) among them are fan-shaped. However, this result may not be a strongly supportive piece of evidence

due to the relatively small sample size of the multipurpose halls. The other trend is that the room shapes have become diverse since the Modern I Period. While one or two room shapes were predominant in the Romantic and Pre-Modern Periods, additional shapes have been employed often since the Modern I Period. The use of arena or semiarena style has gradually increased during the Modern I and Modern II Periods, although the shoebox has still been the most used room shape. Irregular room shapes are regularly observed in these periods. Music performance halls coupled with a surround reverberation chamber represent a noteworthy recent room configuration. These "room-coupled" halls in this study are categorized in the above-mentioned five primary room shapes on the basis of their primary room shape in plan without the reverberation chambers as a modified room shape.

B. Trend in seating capacity

Compared to the Romantic Period (1821–1900), the average seating capacity in the Pre-Modern Period (1901–1950) increased abruptly by approximately 17% (350 seats) from approximately 2070 to 2420 seats. See Table V and Fig. 1(a). In the Modern I Period (1951–1980), the seating capacity continuously increased to approximately 2470 seats but in the Modern II Period (1981–2000), the seating capacity decreased to approximately 2070 seats. During the Modern II Period, the average seating capacity decreased nearly down to that in the Romantic Period and approached that of 23 shoebox halls of all time periods surveyed, which is about 2000. The increase in the two middle periods appears relevant to the increase of fan-shaped or multipurpose fan-

TABLE V. Chronological trends in architectural design indices (average values).

Chronological periods[a]		No. of seats	Room volume (m³)[b]	Volume per seat (m³/seat)	Room dimensions (m)		
					Height	Width	Length
Romantic (1821–1900)	9 halls	2066	16 208	7.84	17.6	22.4	30.7
Pre-Modern (1901–1950)	11 halls	2415	18 156	7.52	18.1	29.1	33.0
Modern I (1951–1980)	23 halls	2471	21 512	8.71	16.3	33.5	35.1
Modern II (1981–2000)	27 halls	2071	21 065	10.17	19.7	29.5	30.6
Overall	71 halls	2326	21 282	9.15	18.2	30.5	32.8
(without Albert Hall and Music Shed)		2244	20 028	8.93	18.0	29.8	32.3

[a]Royal Albert Hall and Music Shed were excluded from the periods of Romantic and Modern I, respectively, because these halls were considered as statistical outliers.
[b]Room volume includes the volume of the main hall and the orchestra enclosure.

FIG. 1. (Color online) Chronological trends in architectural indices with 95% confidence intervals. (a) Seating capacity, (b) room volume, and (c) room volume per seat.

FIG. 2. Chronological trends in room dimensions.

C. Trends in room volume and room volume per seat

The overall trend in room volume generally follows the seating capacity across time periods. See Table V and Fig. 1(b). The average room volume increased from approximately 16200 to 21500 m³ until the Modern I Period. It has stayed the same or decreased slightly since then. The average room volume per seat however was maintained at about 7.7 m³/seat during the Romantic and Pre-Modern Periods. It has tended to increase gradually by 1.35 m³/seat per period since then. See Table V and Fig. 1(c). The value of 10.2 m³/seat reached during the last period (Modern II) is the highest in all time periods.

To summarize prime features observed in the architectural indices of seating capacity, volume and volume per seat during the last chronological period (Modern II) compared to Modern I, the seating capacity decreased by approximately 16% (400 seats) while the room volume decreased by only 2% (450 m³). On the other hand, the room volume per seat increased by approximately 17% (1.5 m³/seat).

D. Trends in room dimensions

Room length and width change similarly over the time periods. See Table V and Fig. 2. The average room length and width increased until the Modern I Period to maximum values of 35.1 and 33.5 m, respectively, and decreased to 30.6 and 29.5 m, respectively, since then. On the contrary, the average room height slightly decreased until the Modern I Period. It increased to about 20 m during the Modern II Period. The increase in width and length observed until the Modern I Period seems related to the increase in the number of large scale fan-shaped halls. In fact, the average values of width and length obtained in the fan-shaped halls are larger by about 7 and 3 m, respectively, in comparison to the overall average values obtained in the 71 halls total. Further, the chronological trend in room volume was found to have a significantly closer relationship with room width than length. See Figs. 1(b) and 2. The statistical correlation coefficient of *adjusted R-square* between room volume and room width is 0.71 at 95% confidence level.

shaped halls during these periods, which hold a larger number of seats at a comparable distance from the stage. This observation is supported by demonstrable increases in average room width from 22.4 to 29.1 m observed in the Pre-Modern Period and up to 33.5 m in the Modern I Period. The average seating capacity of the fan-shaped halls (15 halls) excluding the Music Shed reaches about 2600 seats.

Y. Kwon and G. W. Siebein: Chronological analysis of music halls

(a)

(b)

FIG. 3. (Color online) Average number of rear balcony overhangs. (a) Chronological trend in number of rear balcony overhangs. (b) Average number of rear balcony overhangs versus room shape.

V. CHRONOLOGICAL PHASES IN BALCONY OVERHANG DESIGN INDICES

A. Trend in number of rear balcony overhangs

In the following examination of balcony overlays, simply raised or parterre seating is excluded. The last two centuries (Romantic to Modern II Periods) have seen insignificant change in average number of rear balcony overhangs.
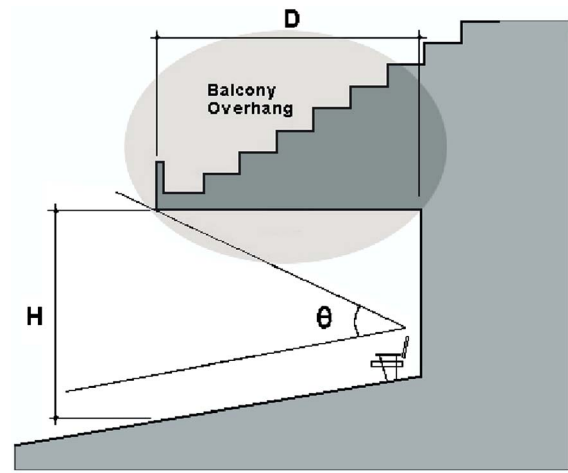
FIG. 4. (Color online) Geometric features of a balcony overhang. D: depth of balcony overhang, H: opening height, and θ: vertical angle of view.

See Fig. 3(a). The average number of rear balcony overhangs in the 71 halls studied here is 1.8. Thus, music performance halls of the sizes examined usually include about two rear balcony overhangs. The number of rear balcony overhangs in the Modern I Period is slightly lower and those in both Romantic and Modern II Periods are slightly higher than the overall average.

The average number of rear balcony overhangs seems to be greatly influenced by room shape. As shown in Fig. 3(b), the arena halls generally have only one rear balcony overhang, while the horseshoe style halls usually have three overhangs. There are one or two overhangs in the fan-shaped halls. For example, as shown in Fig. 3(a), the average number of rear balcony overhangs in the Modern I Period is below the overall average. This resulted from the fact that there are many fan-shaped or arena halls (11 halls out of 23 halls), which have fewer overhangs, but few horseshoe halls are found in this period.

B. Trends in overhang depth (D), opening height (H), and D/H ratio

Geometric features of a generic balcony are annotated in Fig. 4. Over the last two centuries, insignificant changes in the average balcony overhang height (H) measured at the opening aperture occurred. See Fig. 5. However, there have

TABLE VI. Chronological trends in acoustical indices I (average values, unoccupied values, or noted otherwise).

		RT (s) at mid		EDT (s) at mid	BR	G (dB) at mid	ITDG (ms)
Chronological periods[a]		Occupied	Unoccupied				
Romantic (1821–1900)	9 halls	1.81	2.42	2.42	1.14	5.2	20
Pre-modern (1901–1950)	11 halls	1.61	1.91	1.96	1.18	3.8	28
Modern I (1951–1980)	23 halls	1.76	2.10	1.96	1.05	2.8	26
Modern II (1981–2000)	27 halls	1.94	2.30	2.21	1.11	3.3	29
Overall	71 halls	1.82	2.21	2.15	1.11	3.5	26
(Without Albert Hall)		1.81	2.20	2.14	1.11	3.6	26
(Without Albert Hall and Music Shed)		1.81	2.18	2.12	1.11	3.6	27

[a]Royal Albert Hall and Music Shed were excluded from the periods of Romantic and Modern I, respectively.
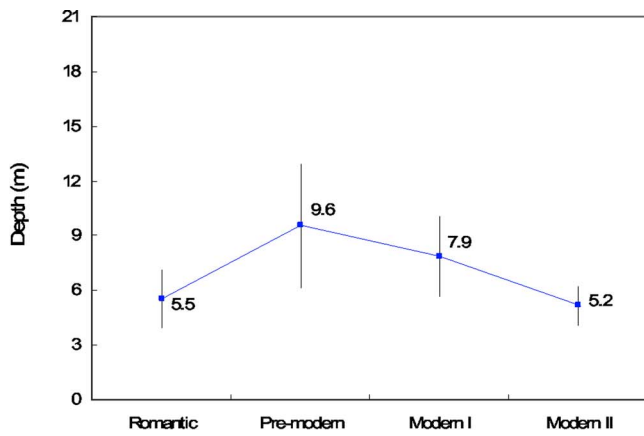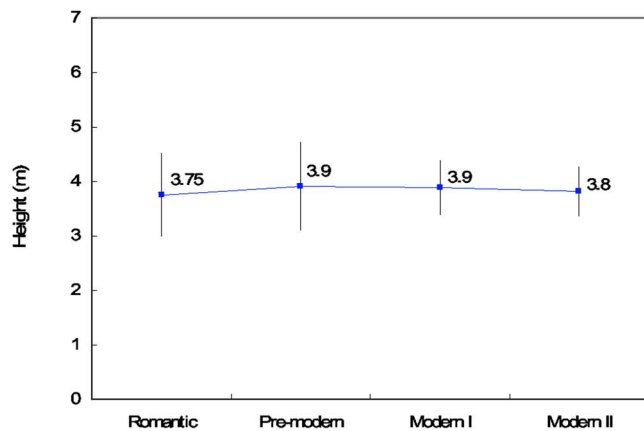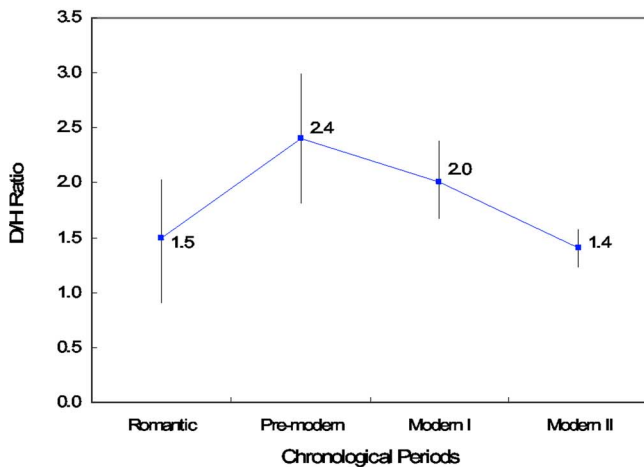
FIG. 5. (Color online) Chronological trends in balcony overhang design indices with 95% confidence intervals. (a) Overhang depth (D), (b) overhang opening height (H), and (c) depth to height (D/H) ratio.

been remarkable changes in average of D/H ratios along with remarkable changes in average overhang depth (D). The values of these two indices—average D and D/H ratio—grew significantly until the Pre-Modern Period, and have noticeably decreased since then. For example, the D/H ratio increased from approximately 1.5 to 2.4 for the halls constructed during the Pre-Modern Period, and decreased to 1.4
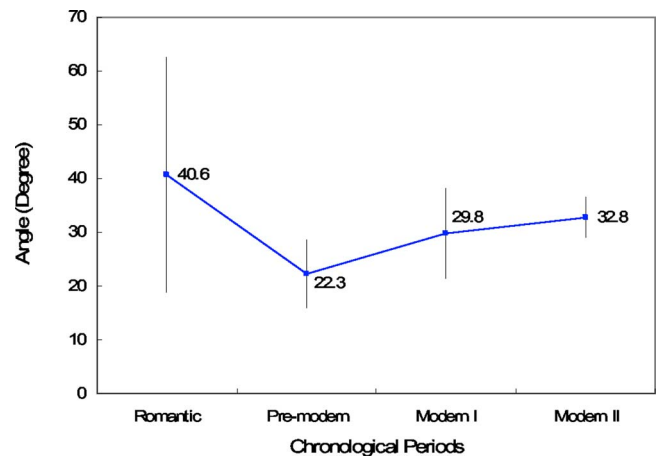
FIG. 6. (Color online) Chronological trend in vertical view angle with 95% confidence intervals.

for the halls opened in the Modern II Period. A primary reason for the former ascending tendency can be inferred from the increase in number of deep balcony overhangs found in fan-shaped halls. In fact, 6 halls (55%) out of 11 halls found in the Pre-Modern Period are fan-shaped, and the average D of these fan-shaped halls reaches 11.1 m. This depth is unusually larger, being deeper by 65% (4.4 m) than the overall average depth of the 71 halls, which is 6.7 m. The increase in number of arena halls (refer to Table IV), which have rather a shallow overhang depth (4.7 m on average), appears reasonable for the latter descending tendency observed both in average D and D/H ratio since the Pre-Modern Period.

C. Trend in vertical angle of view (θ)

As shown in Fig. 6, the trend in vertical angle of view (θ) was found contrary to the trend in overhang depth (D) or D/H ratio discussed earlier. Audience seated at the deepest row under the balcony constructed during the Pre-Modern Period was exposed to a quite narrow vertical view angle of approximately 22°. The average θ has increased to approximately 33° as the D and the D/H ratio have decreased since then.

VI. CHRONOLOGICAL PHASES IN ROOM ACOUSTICAL INDICES

The chronological changes in the room acoustical indices are specified in Tables VI and VII. The acoustical indices examined here are based on the unoccupied data. Otherwise, a description is added in the following discussion—for example, occupied RT.

Trends in RT, EDT, G, and C80. Average values both in EDT and occupied RT decreased to approximately 1.95 and 1.6 s, respectively, from the Romantic through the Pre-Modern Periods but since have gradually increased. See Figs. 7 and 9(a). Those values increased by approximately 0.25 and 0.3 s, respectively, during the Modern II Period compared to the Pre-Modern Period. On the contrary, the C80 trend has an inverse relationship to the EDT or occupied RT trend over periods. The average value of C80 increased

TABLE VII. Chronological trends in acoustical indices II (average values, unoccupied values, or noted otherwise).

Chronological periods[a]		C80 (dB) 500–2000	LF_E 125–1000[b]	IACC 500–2000		BQI (1-$IACC_E$)
				$IACC_E$[c]	$IACC_L$	
Romantic (1821–1900)	9 halls	−2.1	0.19	0.42	0.14	0.58
Pre-modern (1901–1950)	11 halls	0.5	0.16	0.45	0.23	0.55
Modern I (1951–1980)	23 halls	−0.3	0.16	0.53	0.13	0.47
Modern II (1981–2000)	27 halls	−0.9	0.19	0.42	0.15	0.58
Overall	71 halls	−0.7	0.17	0.45	0.15	0.55
(Without Albert Hall)		−0.7	0.17	0.45	0.15	0.55
(Without Albert Hall and Music Shed)		−0.7	0.18	0.44	0.15	0.56

[a]Royal Albert Hall and Music Shed were excluded from the periods of Romantic and Modern I, respectively.
[b]Early lateral energy fraction in 80 ms.
[c]Early interaural cross correlation in 80 ms.

from −2.1 to 0.5 dB during the Pre-Modern Period and has gradually decreased to −0.9 dB since then. See Fig. 9(c). The trend in G value was found to be somewhat different from those observed in the above-mentioned acoustical indices. The average G value decreased by 2.4 dB from 5.2 to 2.8 dB from the Romantic to the Modern I Periods and then increased by 0.5 dB in the Modern II Period. Figure 8(b) shows the trend in G.

VII. CORRELATIONS BETWEEN ARCHITECTURAL DESIGN INDICES AND ROOM ACOUSTICAL INDICES

Fairly robust relationships exist between trends in several architectural indices and acoustical indices across chronological periods. First, the trend in acoustical index of G has a close relationship with room volume or room width. Those relationships are compared in Fig. 8. The G values decreased from the Romantic through Modern I Periods, as the room volume and the room width increased, and the G values have increased since then, as the room volume and width have decreased. The statistical correlation coefficient of *adjusted R-square* at 95% confidence level is 0.87 between G and room volume, and is 0.95 between G and room

FIG. 7. (Color online) Chronological trend in EDT at midfrequencies with 95% confidence intervals.

width. As stated earlier, the *adjusted R-square* between room volume and room width is 0.71.

Second, early IACC ($IACC_E$) follows trends in both room width and room length. However, the above-noted relationship may not be clearly identified in a statistical way due to the limited $IACC_E$ data obtained in each of the time periods.

Third, Fig. 9 compares the trends in room volume per seat with occupied RT and C80. Statistical analysis on this comparison showed that these variables are also closely related to each other. The occupied RT has increased and the C80 has decreased in the 20th Century from the Pre-Modern through the Modern II Periods, as the room volume per seat has increased. The *adjusted R-square* at 95% confidence level is 0.58 between the room volume per seat and the occupied RT averaged through each of the chronological periods.

VIII. CONCLUSIONS

Statistically significant correlations in trends of architectural and acoustical indices in halls for music performance are observed from the 18th through 20th Centuries. Some indices among them are strongly correlated with each other. Note that, from the basic statistical results, large deviations are found among several architectural indices within some of the chronological periods, such as balcony overhang design indices of vertical view angle (θ) and number of balcony overhangs. The average number of rear balcony overhangs in the Modern I Period is 1.6 and its standard deviation is 1.0. Therefore, the chronological classification method used for these architectural design indices of balcony overhangs does not appear to be established. As discussed in the text, these balcony overhang design indices are primarily dependent on room shapes prevalent in each chronological period.

To summarize the statistically significant correlations among architectural and acoustical indices observed across the chronological periods identified, first, the chronological trends between average room volume and average room width were found to be closely correlated with each other. Second, the trend in acoustical index of G has a strong relationship with both room volume and room width. In particu-
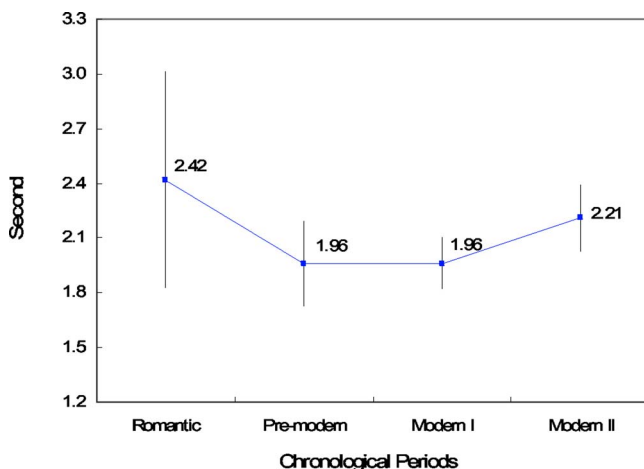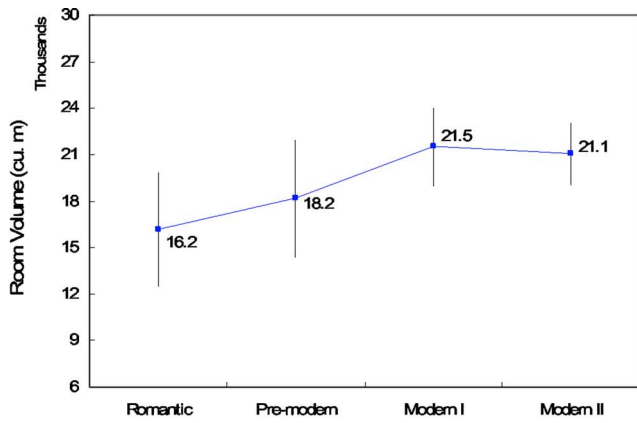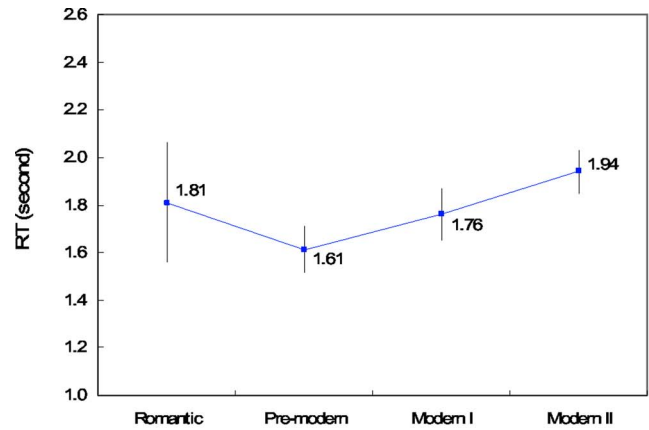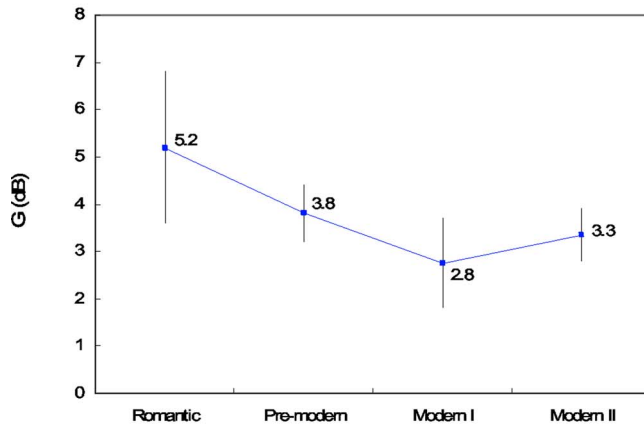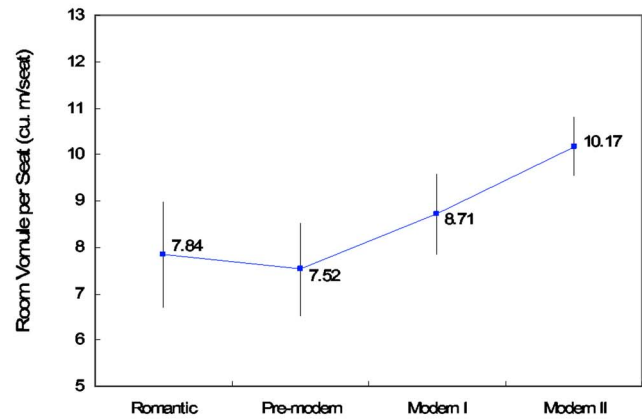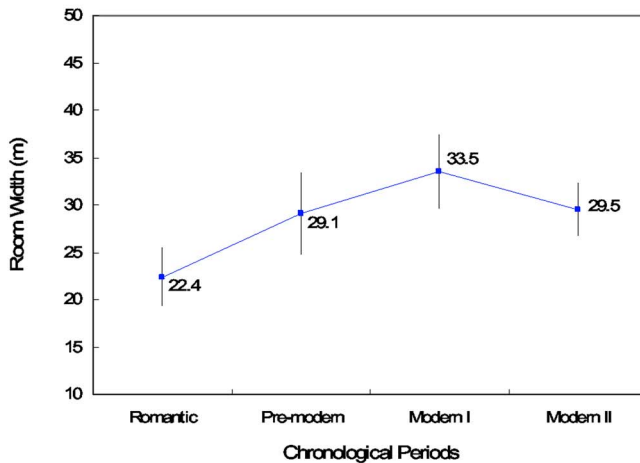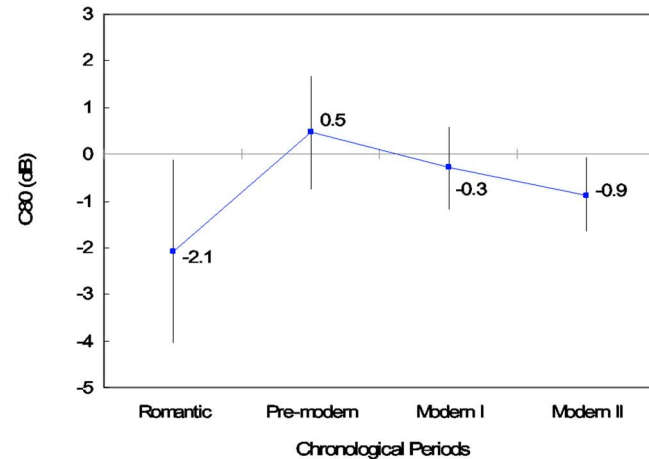
FIG. 8. (Color online) Comparison of the chronological trends of G with room volume and width with 95% confidence intervals. (a) Room volume, (b) G, and (c) room width.

FIG. 9. (Color online) Comparison of the chronological trends of room volume/seat with occupied RT and C80 with 95% confidence intervals. (a) Occupied RT, (b) room volume per seat, and (c) C80.

lar, the statistical correlation coefficient of *adjusted R-square* between G and room width reaches 0.95 at 95% confidence level. Third, the trends observed in acoustical index of either average occupied RT or average C80 were found to be closely related to the trend in average room volume per seat. Last, IACC$_E$ has relationships following room width and length. The above-mentioned relationships argue that some

acoustical indices are strongly dependent on several specific architectural design indices for music performance halls.

Over the past two decade modern period (Modern II: 1981–2000), seating capacity, width, and length have all trended toward reduction even as the room volume has remained relatively constant. This resulted in an increasing room volume per seat ratio. Music performance halls of the

sizes examined generally include two balcony overhangs at the back. The average D and D/H ratio of the first rear balcony overhangs tended to be lower than in prior periods. The average D/H ratio in particular was reduced to approximately 1.4, which is the lowest value over all chronological periods. Notably, acoustical quality has generally improved through the latter half of the 20th Century compared to the former half. The average occupied RT obtained in the past two decades reaches 1.94 s, which is the highest value ever in all periods. The average unoccupied RT and the average EDT are 2.3 and 2.2 s, respectively. These are higher by approximately 0.1 and 0.05 s, respectively, than the overall averages of the 71 halls. The average BQI (1-IACC$_E$) is 0.58, or slightly higher than the overall average of the 71 halls, and the average G and C80 values are 3.3 and −0.9 dB, respectively.

Architectural design development of a performance hall will be continuously linked closely with advancement of room acoustical technologies. Reviewing and understanding of the above-mentioned trends and relationships identified among architectural design indices and room acoustical indices in history should contribute toward acousticians and architects in their new design development.

Beranek, L. L. (**2004**). *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York).

Beranek, L. L. (**1996**). *Concert and Opera Halls: How They Sound* (Acoustical Society of America, Woodbury, NY).

Haan, C., and Fricke, F. R. (**1992**). "Statistical investigation of geometrical parameters for the acoustic design of auditoria," Appl. Acoust. **35**, 105–127.

Hoffman, I. B., Storch, C. A., and Foulkes, T. J. (**2003**). *Halls for Music Performance: Another Two Decades of Experience 1982-2002* (Acoustical Society of America, Woodbury, NY).

Talaske, R. H., Wetherill, E. A., and Cavanaugh, W. J. (**1982**). *Halls for Music Performance: Two Decades of Experience 1962-1982* (Acoustical Society of America, Woodbury, NY).

Measurement of sound power and absorption in reverberation chambers using energy density

David B. Nutter,[a] Timothy W. Leishman, Scott D. Sommerfeldt, and Jonathan D. Blotter
Acoustics Research Group, Department of Physics and Astronomy, Department of Mechanical Engineering, Brigham Young University, Provo, Utah 84602

Reverberation chamber measurements typically rely upon spatially averaged squared pressure for the calculation of sound absorption, sound power, and other acoustic values. While a reverberation chamber can provide an approximately diffuse sound field, variations in sound pressure consistently produce uncertainty in measurement results. This paper explores the benefits of using total energy density or squared particle velocity magnitude (kinetic energy density) instead of squared pressure (potential energy density) for sound absorption and sound power measurements. The approaches are based on methods outlined in current ISO standards. The standards require a sufficient number of source-receiver locations to obtain suitable measurement results. The total and kinetic energy densities exhibit greater spatial uniformity at most frequencies than potential energy density, thus requiring fewer source-receiver positions to produce effective results. Because the total energy density is typically the most uniform of the three quantities at low frequencies, its use could also impact the usable low-frequency ranges of reverberation chambers. In order to employ total and kinetic energy densities for sound absorption measurements, relevant energy-based impulse responses were developed as part of the work for the assessment of sound field decays.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2713667]

I. INTRODUCTION

Current ISO standards governing the measurement of sound power and sound absorption in reverberation chambers rely upon the measurement of spatially averaged squared pressure. The squared pressure at a single point in a chamber is proportional to the potential energy density at that point, representing only a portion of the available energetic information. In 1974, Tichy and Baade suggested that total energy density might be used more efficiently for the determination of sound power, with the expectation that it would have less spatial variation over chamber sound fields.[1] The same year, Cook and Schade used a limited theoretical analysis to conclude that the spatial variance (apparently the normalized spatial standard deviation) of total energy density in a reverberation chamber should be approximately one-half that of potential energy density.[2] They reported experimental results from plane-wave tube measurements and surmised that the spatial variance of total energy density should generally be much less than that of either potential or kinetic energy density. Sepmeyer and Walker described preliminary measurements of total energy density in a reverberation room, confirming that its variation was roughly one-half that of squared pressure.[3] In 1976, Waterhouse and Cook investigated the spatial dependence of the potential, kinetic, and total energy densities for axial, tangential, and oblique modes in reverberation rooms.[4] This extended earlier descriptions of their behaviors in the vicinities of reflecting boundaries.[5,6]

In 1979, Jacobsen used a stochastic diffuse field model with analytical and statistical arguments to show that the normalized spatial variance of the mean-square pressure in a reverberant field should be one.[7] He verified this experimentally for frequencies above the Schroeder cutoff frequency. He then derived the normalized variance for the mean-square value of each Cartesian particle velocity component, finding that they were likewise one, whereas that of the resultant mean-square particle velocity was $\frac{1}{3}$. The normalized variances of the potential, kinetic, and total energy densities were 1, $\frac{1}{3}$, and $\frac{1}{3}$, leading to normalized spatial standard deviations of 1, 0.58, and 0.58, respectively. In 1987, Moryl and Hixson also explored the spatial distribution of the energy densities in reverberation rooms.[8,9] For several pure-tone excitations and a linear traverse in the central region of a reverberation room, they found normalized spatial standard deviations that were 0.94, 0.61, and 0.64, respectively, close to Jacobsen's broader field predictions.

While not directly related to reverberation chamber measurements, several other successful applications of total energy density have been described more recently. One application, originally pursued in three-dimensional enclosures by Parkins *et al.*, used energy density to increase the global extent of active noise control.[10] Their work focused on fields with low modal density. In another application, Bonsi *et al.* proposed a "quadraphonic impulse response" for acoustical enhancement of audio.[11] While this latter work was not known to the authors during the time of research, it is essentially the "energy density impulse response" derived herein for certain reverberation chamber measurements.

Despite the importance of these earlier studies, the use of total energy density for sound power and sound absorption

---
[a]Current affiliation: SARA, Inc., 6300 Gateway Dr., Cypress, CA 90630-4844.

measurements in reverberation chambers has required further development. Because no reliable method was widely available to assess kinetic and total energy densities during the developmental stages of these measurements, spatially averaged squared pressure was accepted for the applications and continues to be used to this day. Nevertheless, recent developments encourage further exploration of the topic.

In very early research during the 1930s, Wolff and Massa used a pressure gradient microphone to estimate particle velocity and compared variations of potential, kinetic, and total energy density in a room.[12] Since that time, several improved methods have been introduced to estimate particle velocity, allowing kinetic and total energy densities to be measured with greater accuracy and consistency.[13–16] The introduction of the Microflown™ sensor, a micromachined device that more directly measures particle velocity, has provided additional means to measure energy density up to 20 kHz.[17,18]

With such feasible measurement options now available, one would anticipate that the suggestion given by Tichy and Baade would be revisited. This was a primary focus of the research reported in this paper. It presents an exploration of their assertion for sound power measurements, while extending the concept through the use of kinetic energy density (or squared particle velocity magnitude) alone. It also explores the use of kinetic and total energy densities for sound absorption measurements. In all cases, it does so by introducing measurement procedures analogous to those found in existing ISO standards (ISO 3741 and 354).[19,20]

Many concerns about sound power measurements in reverberant rooms have been expressed in the past. They address specific problems with the underestimation of sound power at lower frequencies, changes in radiation impedance seen by the sound source, insufficient sound field sampling, and reproducibility at different source positions.[19] The Waterhouse correction is applied to measurements taken in the diffuse field to account for the potential energy stored in interference patterns near reflecting surfaces.[5] Schaffner modified the correction to include absorption effects of the boundaries.[21] Another correction was introduced by Vorländer to account for "missing sound level" using diffuse field equations.[22] Changes in radiation impedance are due to reflections from nearby surfaces and atmospheric conditions. The use of stationary diffusers, a large room volume, and low-frequency absorption can reduce some effects of reflections. A rotating diffuser can also improve spatial averaging of the sound field, which reduces the number of receiver positions required for adequate sampling.[1] Averaging over several source positions may likewise reduce variation effects. Acceptable standard deviation values are outlined in the ISO standards 3741 and 354, and are based on models of statistical distribution of the squared pressure for high modal densities.[23,24]

Reverberation time (T_{60}) measurements used to assess sound absorption also vary with position in a reverberation chamber.[25] Hodgson suggested that exponential decay, while predicted by the Eyring equation[26] for a diffuse field, depends on room shape, room absorption, and the absorption of sound diffusers.[27] Sound may become "trapped" in certain regions of a room, resulting in decays that deviate from theoretical predictions.[28] In practice, the T_{60} may be obtained at a given point by performing backward (Schroeder) integration of the squared-pressure impulse response[29] or by assessing the decay rate after a source producing a steady-state sound field has been turned off. Both the integrated impulse response method and the decay curve method are allowed by ISO 354.

Several interrelated questions served as motivations for the research reported in this paper. First, is it possible to extend the usable low-frequency range of an existing chamber without adding measurement positions or absorption to increase modal overlap? If so, reverberation chambers could service several types of measurements with fewer required modifications. Second, could smaller reverberant rooms that fail to meet volume requirements or other rooms that otherwise fail to satisfy qualification standards be used to obtain reliable measurement results, thus enhancing means for those without access to ideal facilities? Third, if measurements were taken of an acoustic field quantity with greater spatial uniformity than squared pressure, how would this reduce the required number of measurement positions in a given chamber? Fourth, could currently allowed deviation values for reverberation chamber measurements be substantially reduced through alternative field measurements? The research was undertaken to address these questions through specific exploration of kinetic and total energy densities in reverberant rooms.[30] While the work involved some theoretical and numerical modeling, it primarily involved experimental efforts.

The following sections address pertinent concepts of the energy density methods and calculations required to evaluate sound absorption and sound power in reverberation chambers. They include derivations of pertinent energy-based impulse responses. Experimental results for two chambers are presented, along with further discussion of the measurement implications.

II. METHODS

Sound absorption and sound power measurements were taken in two rooms. The first was a qualified reverberation chamber with dimensions of $4.96\text{m} \times 5.89\text{m} \times 6.98\text{m}$, a volume of approximately 204 m^3, and a Schroeder frequency of 410 Hz (with no low-frequency absorption in place). The second was a much smaller chamber with dimensions of $5.70\text{m} \times 4.30\text{m} \times 2.50\text{m}$, a volume of approximately 61 m^3, and a Schroeder frequency of 552 Hz (with no low-frequency absorption in place). Both chambers incorporated stationary diffusers that were consistently used throughout the investigation. Because the volume of the smaller chamber did not meet qualification standards, its conventional low-frequency measurements were expected to suffer in accuracy.

A Microflown™ sensor, consisting of one acoustic pressure sensor and three orthogonal particle velocity sensors for each Cartesian component, was used to measure the chamber sound fields at discrete points. The output signal of each transducer was fed to a signal conditioning box. The corresponding outputs of the box were then fed to a control room

for digitization, analysis, storage, and later processing. A dodecahedron loudspeaker was used as a source to assess the sound absorption of test specimens. A TEF 20 analyzer drove the loudspeaker amplifier with a maximum-length sequence and sequentially monitored each output channel of the sensor to assess the associated room impulse responses for the selected source-receiver locations. A Matlab routine was developed to compute the Schroeder decay curves using the integrated impulse response method. A Sony DAT recorder was used to record the steady-state responses from a Brüel & Kjær reference sound source for sound power measurements. These measurements were processed using another Matlab routine. In all cases, the temperature, relative humidity, and barometric pressure of the chambers were monitored at positions close to the sensor. The following sections provide additional details about the specific methods used for the sound absorption and sound power measurements.

A. Sound absorption

1. Standard formulation for equivalent absorption area

From ISO 354, the equivalent sound absorption area of a test specimen for a given frequency band is

$$A_T = 55.3V \left(\frac{1}{c_2 T_2} - \frac{1}{c_1 T_1} \right) - 4V(m_2 - m_1), \tag{1}$$

where T denotes the arithmetic mean value of T_{60}, c is the speed of sound, V is the room volume, and m is the power attenuation coefficient. The subscripts 1 and 2 refer to values of the chamber when empty and with the test specimen present, respectively. The value for m can be obtained from the relationship

$$m = \frac{\alpha}{10 \log(e)}, \tag{2}$$

where α is the pure-tone sound-attenuation coefficient for atmospheric absorption, in decibels per meter. This coefficient is obtained by inserting the values of temperature, relative humidity, and barometric pressure into equations detailed in ISO 9613.[31] The accuracy of the pure-tone attenuation coefficient is estimated to be ±10% under recommended atmospheric measurement conditions.[20] This has strong effects on measurements at higher frequencies, as will be discussed later.

For this research, two different types of samples were measured. The first was a planar 5-cm-thick semi-rigid fiberglass insulation panel resting on the floor with an A-mounting (see ISO 354). The total exposed surface area of the material, including edges, was 10.65 m^2 for the large chamber and 6.11 m^2 for the small chamber. The second sample consisted of three upholstered office chairs. The results for the latter are presented elsewhere.[30]

2. Use of alternate impulse responses

The behavior of a linear time-invariant system may be characterized by its impulse response $h(t)$, defined as the time-domain output of the system due to the application of a unit impulse function (Dirac delta function) $\delta(t)$ at the input,

a time τ beforehand. For an arbitrary input function $a(t)$, the output $b(t)$ is given by the convolution integral

$$b(t) = \int_{-\infty}^{\infty} h(\tau) a(t - \tau) d\tau, \tag{3}$$

such that if $a(t) = \delta(t)$,

$$b(t) = \int_{-\infty}^{\infty} h(\tau) \delta(t - \tau) d\tau = h(t). \tag{4}$$

In an enclosed sound field, the impulse response $h_p(\tau)$ between a source and a receiver is typically considered in terms of the acoustic pressure $p(t)$ at the receiver "output" position. However, the individual Cartesian components of the particle velocity $[u_x(t), u_y(t), \text{and } u_z(t)]$ also satisfy the linear wave equation[32] and can therefore be used for the definition of additional impulse responses. For the x component of the particle velocity, the convolution integral becomes

$$u_x(t) = \int_{-\infty}^{\infty} h_{u_x}(\tau) a(t - \tau) d\tau, \tag{5}$$

where the impulse response $h_{u_x}(\tau)$ is defined by the application of the unit impulse:

$$u_x(t) = \int_{-\infty}^{\infty} h_{u_x}(\tau) \delta(t - \tau) d\tau = h_{u_x}(t). \tag{6}$$

Similar relationships follow for the y and z components of the particle velocity. Since the square of the vector particle velocity magnitude is defined as

$$u^2(t) = |\mathbf{u}(t)|^2 = u_x^2(t) + u_y^2(t) + u_z^2(t), \tag{7}$$

one can analogously define a squared impulse response for the vector particle velocity magnitude:

$$h_u^2(t) = h_{u_x}^2(t) + h_{u_y}^2(t) + h_{u_z}^2(t). \tag{8}$$

An impulse response $h_w(\tau)$ associated with total energy density then follows with appropriate weightings:

$$h_w(t) = \frac{1}{2\rho_0 c^2} h_p^2(t) + \frac{\rho_0}{2} h_u^2(t). \tag{9}$$

To study the impact of each specified measurements, the impulse responses obtained in Eqs. (8) and (9) can be used in calculations in place of $h_p^2(t)$.

Because absorption data are typically observed in one-third-octave frequency bands, the primary impulse responses are filtered before they are squared in order to preserve spectral information. Figure 1 shows examples of squared pressure, squared velocity magnitude, and total energy density impulse responses in the 250-Hz one-third-octave band at one position in the larger chamber. The impulse responses are displayed on logarithmic amplitude scales. Of practical concern is the error introduced through the factors ρ_0 and c. In many cases, the relative amplitude of the squared impulse response is not important, since T_{60} is merely derived from the slope of the integration curve. However, the combination of several impulse responses weighted by these constants suggests that errors of the measured atmospheric values
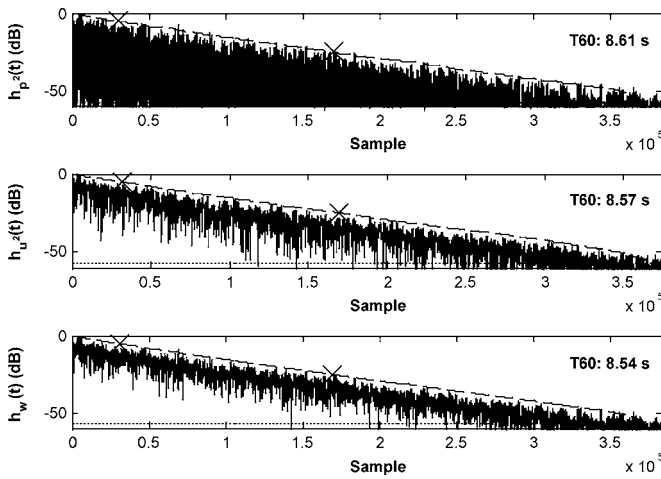
FIG. 1. Impulse responses of pressure, particle velocity magnitude, and total energy density in the 250-Hz one-third-octave band, obtained at one position in the large chamber. The abscissa corresponds to the sample number for a 48 kHz sampling frequency. The dashed line is the Schroeder integration curve. The horizontal dotted line shows the noise floor. The integration starts where this line intersects with the Schroeder curve. The cross marks show the −5 and −25 dB down points.

could lead to errors in the total energy density impulse response. The significance of these errors will be treated in a later section.

B. Sound power

1. Standard formulation using sound pressure

The sound pressure level (SPL or L_p) at a point in a chamber is given by

$$L_p = 10 \log\left(\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2}\right), \tag{10}$$

where p_{rms} is the measured rms acoustic pressure and p_{ref} is the reference acoustic pressure (20 μPa). Following procedures given in ISO 3741, the sound power level (PWL or L_W) of a source under test is given by

$$L_W = \bar{L}_p + \left[10 \log \frac{A}{A_0} + 4.34 \frac{A}{S} + 10 \log\left(1 + \frac{Sc}{8Vf}\right) \right.$$
$$\left. - 25 \log\left(\frac{427}{400} \sqrt{\frac{273}{273 + T_c}} \frac{B}{B_0}\right) - 6 \right], \tag{11}$$

where $\bar{L}_p$ is the SPL in a given frequency band, averaged over all source and microphone positions, A is the equivalent sound absorption area in the test room, A_0 is 1 m², S is the total surface area of all room surfaces in m²,[33] V is the room volume in m³, f is the mid-band frequency of the measurement in Hz, c is the speed of sound in m/s at temperature T_c (in °C), B is the atmospheric pressure in Pa, and B_0 is the reference atmospheric pressure, 1.013×10^5 Pa. The first and last terms in the square brackets follow from the relationship between sound power level and sound pressure level in a diffuse field. The second term was introduced by Vorländer to compensate for an underestimation of the sound field using Sabine's equation. The third term is the Waterhouse correction. The fourth term accounts for at-

mospheric effects in the room and corrects the value to that measured with a characteristic specific acoustic impedance of 400 Pa·s/m.

2. Formulation using other field quantities

Two alternate quantities were considered in this study to obtain the sound power level of a source under test. One was the sound velocity level (SVL or L_u). The other was the total sound energy density level (SEDL or L_w). The sound velocity level is defined as ten times the base 10 logarithm of the ratio of the time-mean-square particle velocity of a given sound or vibration to the square of a specified reference particle velocity:[34]

$$L_u = 10 \log\left(\frac{u_{\text{rms}}^2}{u_{\text{ref}}^2}\right). \tag{12}$$

The reference value of u_{ref} has not been clearly standardized. Clause A.1 of ANSI S18-1989 notes that a reference particle velocity of 10 nanometers per second (10 nm/s) is used in ANSI S1.8-1969. This is in contrast with the 1 nm/s preferred by ISO 1683-1983.

Plane-wave conditions describe the relationship between particle velocity and acoustic pressure as

$$u = \frac{p}{\rho_0 c}. \tag{13}$$

It is reasonable then to assume that the reference velocity might be related to the reference pressure in the same way:

$$u_{\text{ref}} = \frac{p_{\text{ref}}}{\rho_0 c}. \tag{14}$$

However, if the characteristic impedance $\rho_0 c$ is estimated to be 400 Pa·s/m and $p_{\text{ref}} = 20$ μPa, the reference velocity would be 50 nm/s. This value is much greater than either the ANSI or ISO suggestion, but its use is not uncommon.[35] In any case, if p_{ref} is considered to be the primary standard, a fixed u_{ref} may not remain sufficiently accurate, since ambient atmospheric conditions determine the value of $\rho_0 c$.

The total sound energy density level may be expressed as

$$L_w = 10 \log\left(\frac{\langle w \rangle_t}{w_{\text{ref}}}\right), \tag{15}$$

where $\langle w \rangle_t$ is the time-averaged total energy density and w_{ref} is the reference energy density. The reference value may be determined from a summation of the squared pressure and particle velocity reference values, with appropriate weightings:

$$w_{\text{ref}} = \frac{1}{2\rho_0 c^2} p_{\text{ref}}^2 + \frac{\rho_0}{2} u_{\text{ref}}^2. \tag{16}$$

For a plane wave, the sound pressure level, sound velocity level, and sound energy density level should all be equal, as long as the appropriate reference values are used. Thus, once L_u and L_w have been calculated, the sound power level can be determined from either of these values by inserting them
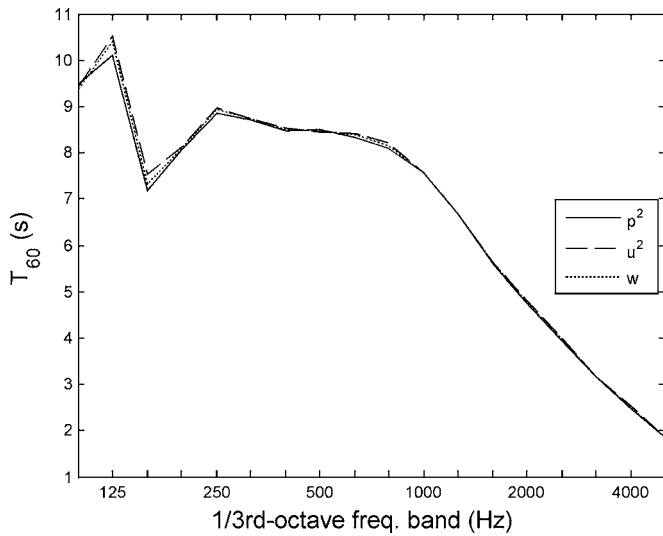
FIG. 2. Average T_{60} measurements based on squared pressure, squared velocity magnitude, and total energy density in the empty large chamber.
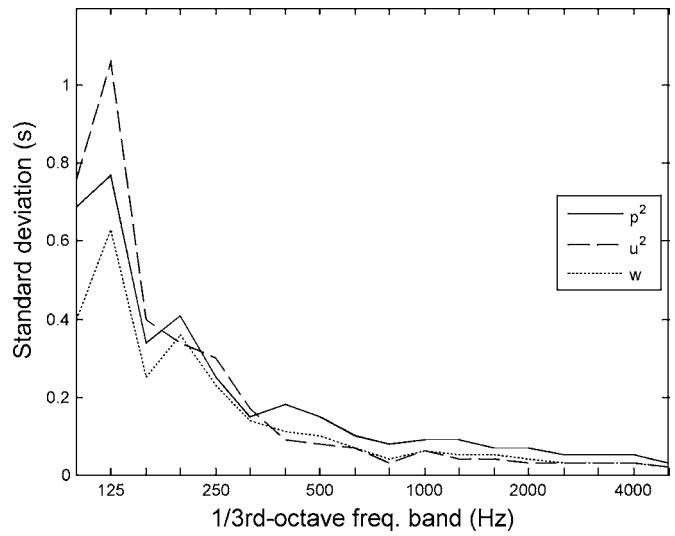
FIG. 3. Standard deviation of T_{60} measurements based on squared pressure, squared velocity magnitude, and total energy density in the empty large chamber.

in place of L_p in Eq. (11), with appropriate attention to correction terms. For this study, the Waterhouse correction term was chosen to be consistent when L_u was used because the value of the integrated interference pattern for kinetic energy density is the same as that for potential energy density.[5] When L_w was used, the number 8 in the denominator of the term was replaced by 4, for an assumed combination of the potential and kinetic energy density interference patterns. All other correction terms remained the same. The actual levels obtained in an enclosure are contingent upon the diffuseness of the sound field. Because the variation in the sound field is different for each quantity, the average sound levels will also differ for a given number of source-receiver locations.

III. EXPERIMENTAL RESULTS

A. Sound absorption

1. Reverberation time

For all T_{60} measurements, 12 source-receiver positions were used in each room for averaging, as required by the ISO 354 standard. The average T_{60} values were obtained from impulse responses based on squared pressure (p^2), squared particle velocity magnitude (u^2), and total energy density (w). They were then examined and compared. The probability of one measurement quantity outperforming the others (in terms of spatial uniformity) was obtained for various cases. These methods were repeated with and without the inclusion of test materials in the rooms.

Figure 2 shows the average T_{60} values and Fig. 3 shows their standard deviations for the empty large chamber using all 12 measurement positions. The average values agree, particularly at higher frequencies. Minor differences occur at 125 and 160 Hz. Total energy density w produces a notably lower standard deviation than p^2 at nearly all frequencies. Based on previous energy density studies, this was an anticipated result. At higher frequencies, u^2 significantly outperforms p^2. In fact, it appears to perform as well as, if not

better than, w at these frequencies. This phenomenon was repeated in later observations. At lower frequencies, p^2 performs better than u^2.

To further investigate and visualize the performance of u^2 and w, probability tests were conducted to analyze the minimum number of positions required to achieve results as good as those for 12 standard measurement positions using p^2. All possible combinations of the data were evaluated as a function of the number of positions. The probability P that the standard deviation of T_{60} from r randomly chosen u^2 positions is lower than the standard deviation of T_{60} from all 12 p^2 positions is shown in Fig. 4. The probability for r randomly chosen w positions is shown in Fig. 5. Of the two quantities, w generally performs best over all frequencies. At
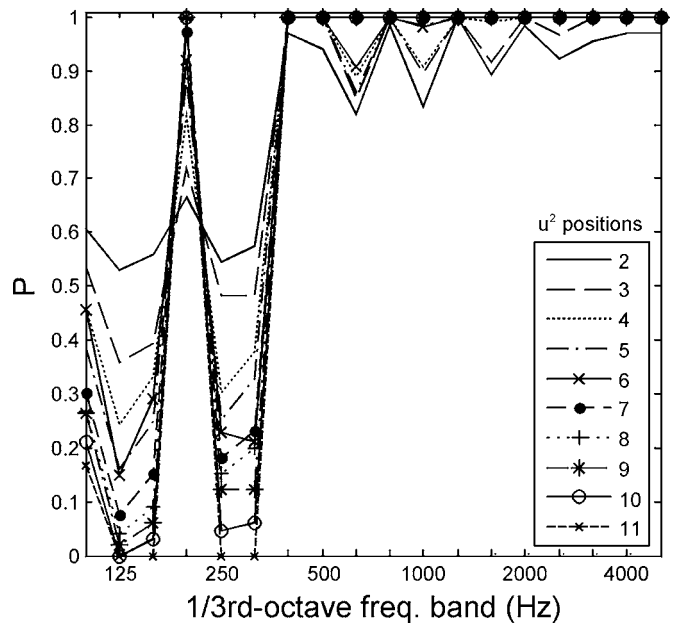
FIG. 4. Progressive probabilities that a given number of positions of squared velocity magnitude (u^2) outperform 12 squared pressure (p^2) measurements in the empty chamber.

Nutter *et al.*: Energy density methods in reverberation chambers
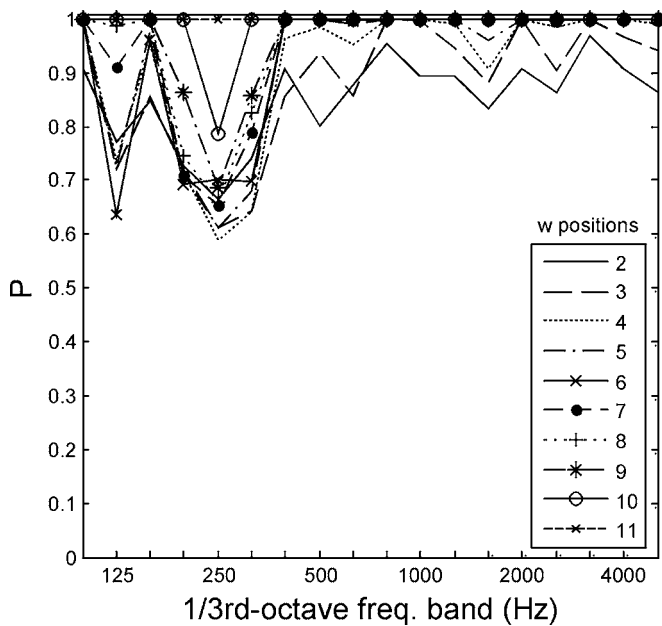
FIG. 5. Progressive probabilities that a given number of positions of total energy density (w) outperform 12 squared pressure (p^2) measurements in the empty chamber.

FIG. 6. Progressive probabilities of lower variation in u^2 vs. p^2 in the empty large chamber.

higher frequencies, six w positions ensure a lower standard deviation than the 12 p^2 positions. Standard deviation values for u^2 also do very well at higher frequencies; seven positions are enough to outperform the p^2 positions. In both cases, a transition to lower probabilities is apparent below the Schroeder frequency of 410 Hz. While the number of measurement positions needed below the Schroeder frequency becomes greater, one generally observes that fewer w positions are necessary than either p^2 or u^2 positions. Above the Schroeder frequency, either w or u^2 leads to fewer measurement positions.

For yet another point of view, an additional comparison was made by evaluating the probability that any number of randomly chosen positions for one quantity led to a lower variation than the same number of randomly chosen positions for any other quantity. Two relationships were plotted: u^2 vs. p^2, and w versus p^2. The results are given in Figs. 6 and 7, respectively. The probability that variation is lower for either u^2 or w increases with the number of positions. To see if the performance of u^2 versus w was indeed significant, an F test of unequal variances was used.[36] The test revealed that for a 95% confidence level, the standard deviation values were statistically equal, except at 100 Hz, where that for w was smaller. This phenomenon was further investigated analytically, as will be discussed in the next section.

The T_{60} results were also compared when absorptive materials were present in the room. The measurements and comparisons were likewise repeated for the small chamber. As reported in Ref. 30, the results for the various averages and standard deviations were similar to those discussed above, especially at frequencies above the Schroeder frequency.

2. Absorption coefficients

The sound absorption coefficients for the test samples were calculated according to the procedures outlined in ISO
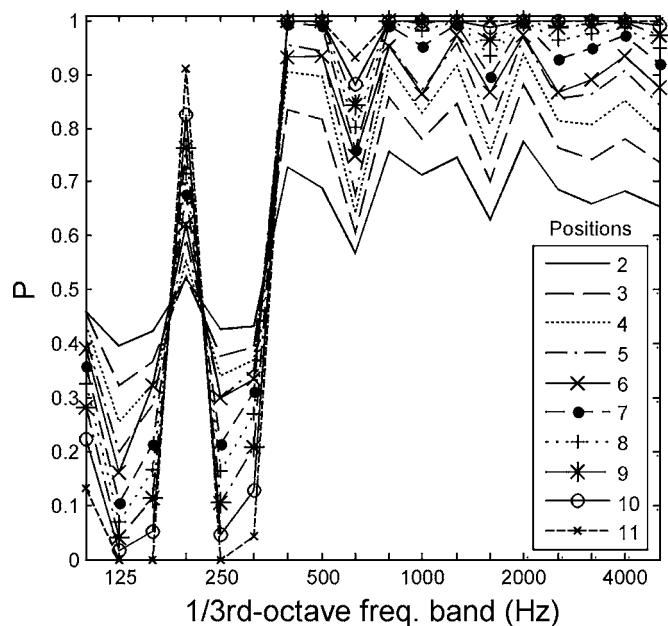
354. The calculations were based on the impulse responses from the 12 source-receiver positions. From the T_{60} results, half the w positions were sufficient to obtain results comparable with the p^2 results above the Schroeder frequency. Six u^2 and w positions were thus randomly chosen and compared to all 12 p^2 positions for calculation of the absorption coefficients. Although this number was not sufficient for the entire measurement bandwidth, it showed that the results are generally similar to those of the conventional measurements. Figures 8 and 9 show the absorption coefficient (α_s) results obtained from the two chambers. The ±10% accuracy window described in ISO 9613-1 was calculated based on the atmospheric absorption α obtained for the squared-pressure
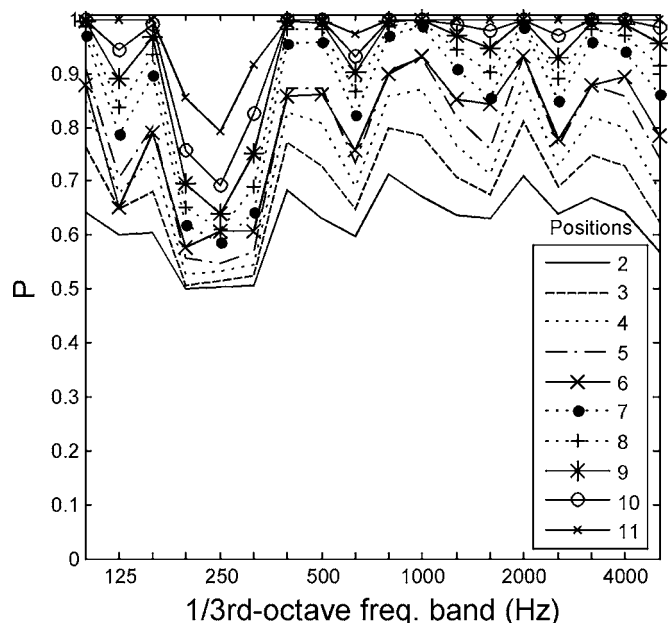
FIG. 7. Progressive probabilities of lower variation in w vs. p^2 in the empty large chamber.
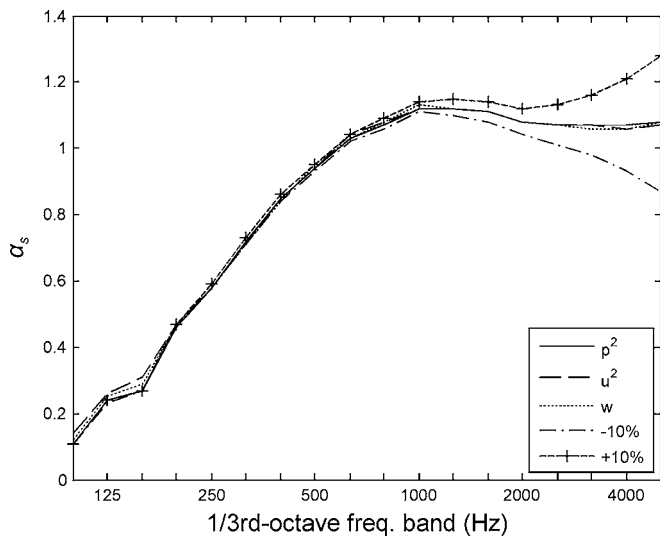
FIG. 8. Measured absorption coefficients of 5 cm thick fiberglass insulation in the large chamber. The three curves are obtained from 12 p^2 positions and from 6 random u^2 and w positions. The accuracy limits due to calculation of the air attenuation coefficient are also plotted.

FIG. 10. Sound power levels (PWL) for a reference sound source at one position in the large chamber. The three curves labeled PWL_p, PWL_u, and PWL_w were obtained from six spatially averaged SPL, SVL, and SEDL measurements, respectively.

measurements. (In other words, the window is calculated from the values 1.1α and 0.9α at each frequency.)

These plots demonstrate that the results are very similar for a given room, no matter which measurement quantity is used. Because of the smaller spatial variations of T_{60} for both u^2 and w, fewer measurement positions can be used to obtain the absorption coefficients. In this case, the number of positions was reduced by one-half. A comparison of the measurements from the two chambers reveals that they are quite similar, although some differences do exist.

B. Sound power

A Brüel and Kjær 4204 reference sound source was used for all sound power measurements. Six receiver positions and one source position were chosen in the large chamber,

the minimum required by the ISO 3741 standard. Three receiver positions and two source positions were chosen in the small chamber, due to its smaller working area. The spatially averaged SPL, SVL, and SEDL values were obtained and used to determine the sound power of the source. Figure 10 shows the results from the large chamber and Fig. 11 shows the corresponding standard deviations. Similar results are given for the small chamber in Figs. 12 and 13 (note that Fig. 13 uses a larger vertical scale than Fig. 11). The allowed standard deviations of sound pressure levels for the measured field points (see ISO 3741) are included in the standard deviation plots. Where the measured standard deviations exceed the limits, more source and/or receiver positions become necessary.

FIG. 9. Measured absorption coefficients of 5 cm thick fiberglass insulation in the small chamber. The three curves are obtained from 12 p^2 positions and from 6 random u^2 and w positions. The accuracy limits due to calculation of the air attenuation coefficient are also plotted.
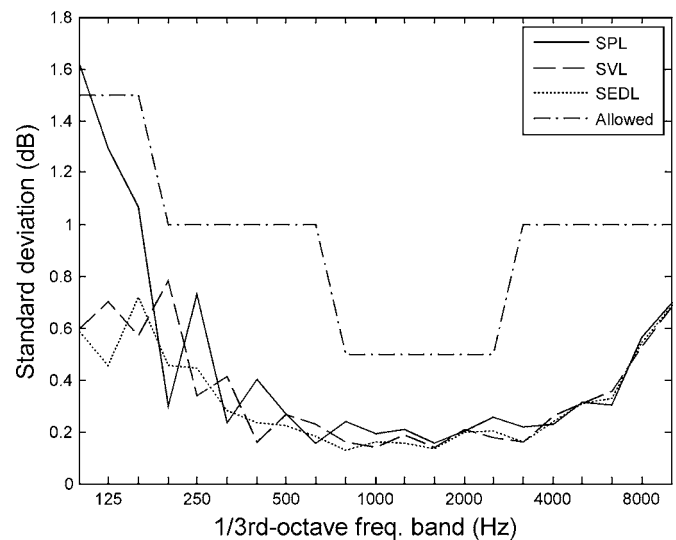
FIG. 11. Measured and allowed standard deviations of sampled field quantities used to assess the sound power produced by a source in the large chamber. Standard deviation values exceeding the allowed limits require additional source and/or receiver positions. The SVL and SEDL measurements are shown to extend the usable low-frequency range of the chamber without the addition of low-frequency absorption.

Nutter *et al.*: Energy density methods in reverberation chambers
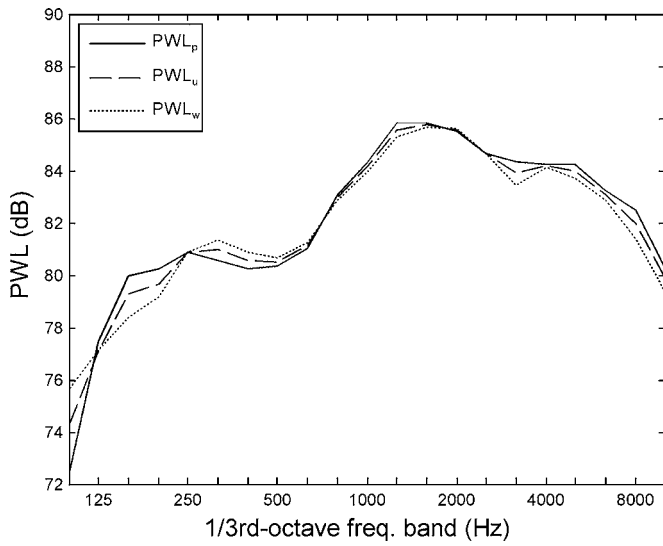
FIG. 12. Sound power levels (PWL) for a reference source in the small chamber. The three curves labeled PWL_p, PWL_u, and PWL_w were obtained from six spatially averaged SPL, SVL, and SEDL measurements, respectively, which involved two source positions.

The sound power levels obtained from the SPL, SVL, and SEDL measurements are similar within each chamber, while the results from one chamber vary slightly from those of the other. In the large chamber, the standard deviation of the SPL values barely exceeds the allowed deviation at the lowest frequencies. As suggested before, this problem is due to low modal overlap and would likely be corrected with the addition of low-frequency absorbers. Standard deviation values determined from SVL and SEDL perform much better at these lower frequencies. This suggests that the addition of low-frequency absorbers may not be necessary if sound power is determined via particle velocity magnitude or total energy density.

As expected, the results for the small chamber show that
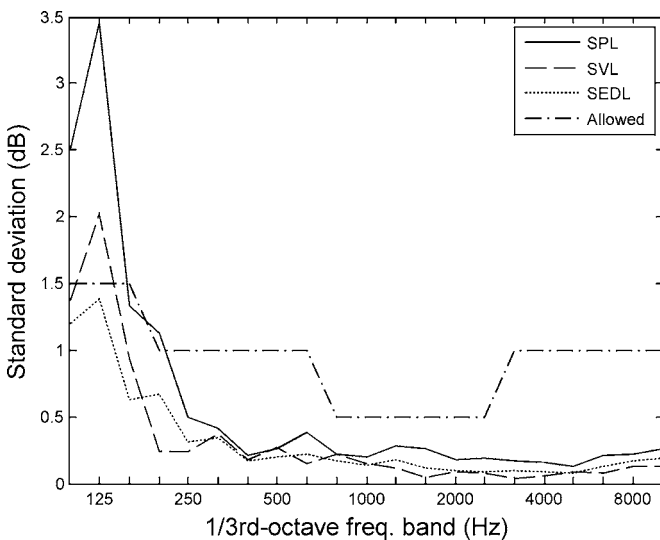
FIG. 13. Measured and allowed standard deviations of sampled field quantities used to assess the sound power produced by a source in the small chamber. Standard deviation values exceeding the allowed limits require additional source and/or receiver positions. The SEDL measurements are shown to extend the usable low-frequency range of the chamber.

the standard deviation of the SPL is very poor at lower frequencies. The standard deviation of the SVL is better, but it still exceeds the allowed value in the 125-Hz octave band. However, the standard deviation of the SEDL is well within allowed limits at all frequencies, including the problematic lower frequencies. This suggests that total energy density measurements could be used to better estimate low-frequency sound power of sources in smaller rooms with modal overlap that is typically considered inadequate.

IV. ADDITIONAL DISCUSSION

A. Theoretical and numerical modeling

Simple theoretical and numerical models were developed to simulate the acoustical characteristics of the large chamber and enhance understanding of the experimental measurement results.[30] The room was modeled as an ideal rectangular enclosure with dimensions $5m \times 6m \times 7m$, large wall impedance, and spectrally uniform damping. The frequency-dependent complex acoustic pressure amplitude $\hat{p}(x,y,z)$ was calculated at points in the enclosure due to a point source at a fixed position using a familiar eigenfunction expansion.[37] The Cartesian components of the frequency-dependent particle velocity $\hat{\mathbf{u}}(x,y,z)$ were determined from the expansion by applying Euler's equation for each dimension in its time-harmonic form. The time-averaged potential energy density at each point was subsequently obtained using the expression

$$\langle w_p(x,y,z)\rangle_t = \frac{1}{4\rho_0 c^2}\hat{p}(x,y,z)\hat{p}^*(x,y,z) = \frac{1}{4\rho_0 c^2}|\hat{p}(x,y,z)|^2 \tag{17}$$

and the time-averaged kinetic energy density was obtained using the expression

$$\langle w_k(x,y,z)\rangle_t = \frac{\rho_0}{4}\hat{\mathbf{u}}(x,y,z)\cdot\hat{\mathbf{u}}^*(x,y,z)$$

$$= \frac{\rho_0}{4}[|\hat{u}_x(x,y,z)|^2 + |\hat{u}_y(x,y,z)|^2$$

$$+ |\hat{u}_z(x,y,z)|^2]. \tag{18}$$

The time-averaged total energy density was then calculated as the sum of the potential and kinetic energy density components.

Thousands of receiver locations were required in the calculations to adequately describe the spatial variation of the three-dimensional field. While it would have been difficult and time consuming to perform actual measurements at such high resolution throughout the chamber volume, it was relatively simple to simulate them using the computer model. An effort was made to include a sufficient number of modes in the calculations. For an analysis through the 5-kHz one-third-octave band, up to 7.5 million modes were involved. An absorption coefficient of 0.02 was chosen for all room surfaces, at all frequencies of interest. This value produced a T_{60} value of 7.9 s, a uniform modal bandwidth of 0.27 Hz,[37] and a Schroeder frequency of 388 Hz. The modal damping factor δ_N also followed from the coefficient.[38]
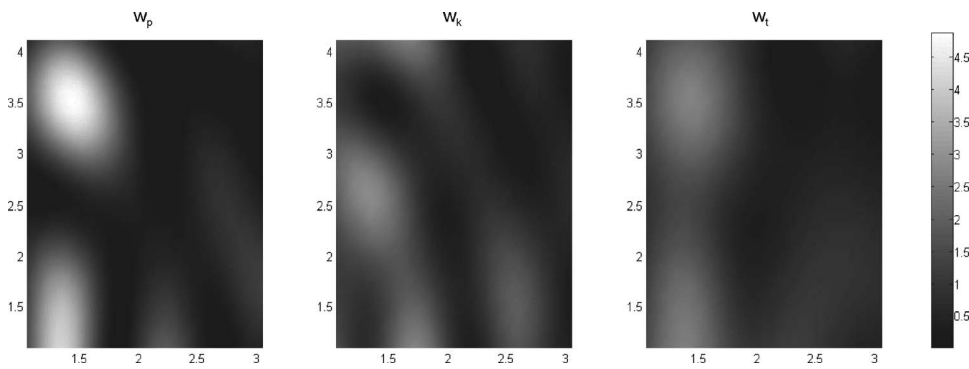
FIG. 14. Mapped potential, kinetic, and total energy density fields for the $z=3$ m plane of the modeled large chamber, in the 125-Hz one-third-octave band. For this example, the source was positioned in the lower corner of the chamber at $(x, y, z) = (0, 0, 0)$.

The spatial variation was calculated for the one-third-octave bands between 100 and 5000 Hz. The calculations were obtained by summing the contributions of 1-Hz bins within each band. To reduce computation time, only the modes that fell within nearly ten times the modal bandwidth at each frequency bin were used, since those outside this range contributed much less significantly to the results. The source strength of the point source was set to unity. In the case illustrated here, the source was positioned in a lower corner of the room ($x_0=0$ m, $y_0=0$ m, $z_0=0$ m). Field quantities were calculated at 0.1-m increments in x, y, and z, beginning 1 m away from each wall, and 2 m away from the source, yielding 52, 111 positions in total. The distances from the walls and source were chosen for consistency with the practices outlined in ISO 3741.

Pertinent field values for each plane in z were saved to separate Matlab files for later extraction and analysis. Each file included pressure, particle velocity magnitude, potential energy density, kinetic energy density, and total energy density. The data were eventually grouped for two cases: variation over the entire volume and variation over each plane in z. Given the fixed room properties, the variables affecting the outcome included the source position, receiver plane, and frequency. To equitably compare the variation for each field quantity, the standard deviation was calculated after normalizing the data by its mean value. This resulted in a new average value of 1 and a dimensionless standard deviation. This method of normalization was chosen because measurements deal with average values.

Figure 14 shows the normalized potential, kinetic, and total energy density fields for the 125-Hz one-third-octave band and the $z=3$ m plane of the model. Similar results could be presented for any plane in the room. One notes from the figure that the maxima of the potential energy density field often fall near the minima of the kinetic energy density field and vice versa. Figure 15 shows the frequency-dependent standard deviations of the normalized fields throughout the entire volume, with the source in the corner as indicated above. Just as for the experimental sound power measurements for the large chamber, both kinetic and total energy density outperformed potential energy density in terms of spatial uniformity. The performances of kinetic and total energy density are quite similar. At higher frequencies, the results agree reasonably well with the predictions given by Jacobsen.[7] An F test for unequal variances revealed that in most cases, for a 99% confidence level, the standard deviations of kinetic and total energy densities are statistically equal.

B. Uncertainty errors due to atmospheric conditions

As explained earlier, the uncertainties in reverberation chamber measurements come from many sources. Some involve the atmospheric conditions in the rooms. For the acoustical measurements reported in this paper, temperature, relative humidity, and barometric pressure were regularly evaluated. The respective values had the following margins of error for the measurement instrument used: ±1 °C, ±7% RH, and ±700 Pa. All three values were used to obtain the air attenuation coefficient α. The coefficient is based on a theoretical calculation of the relaxation frequencies of nitrogen and oxygen, as detailed in ISO 9613, with accuracy to ±10%. Of the three atmospheric values, temperature most strongly affects the speed of sound. An uncertainty of ±1 °C results in a variation of ±0.6 m/s.

The uncertainty present in the air attenuation coefficient has a considerable effect on the equivalent absorption area at frequencies above 1 kHz. For example, the error value of α at 5 kHz for the fiberglass insulation absorption in the large chamber is ±0.21, according to the 10% accuracy limits. In comparison, the error value of α due to an uncertainty of
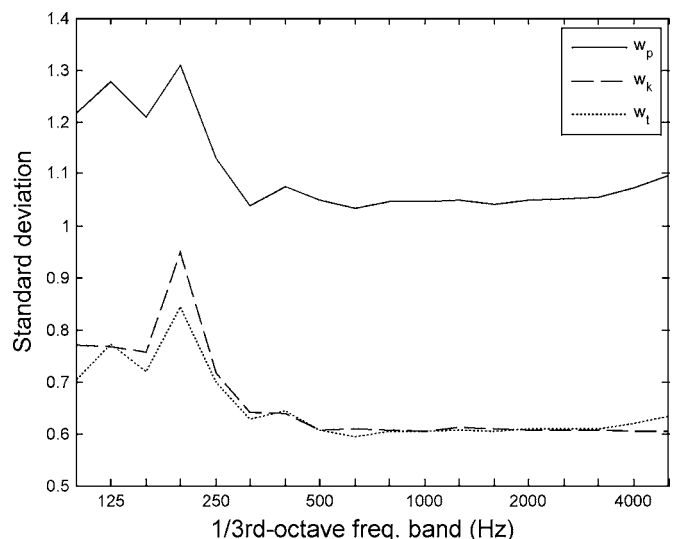
FIG. 15. Standard deviation curves for the normalized potential, kinetic, and total energy density throughout the modeled room volume.

Nutter *et al.*: Energy density methods in reverberation chambers

±1 °C alone is ±0.08. Accordingly, it appears that the uncertainty from the measured atmospheric values is not as significant as that inherent in the standard itself. However, the uncertainty effects are further accentuated in the calculation of energy density, which uses both ρ_0 and c. The uncertainties of the three atmospheric values result in an overall uncertainty of ±0.013 kg/m^3 for ρ_0. Thus, the energy density impulse response might introduce additional errors in the measured T_{60}, an issue absent in the squared-pressure impulse response. This also affects the sound power level obtained from the SEDL by ±0.1 dB. Are these effects negligible? Since T_{60}, absorption coefficients, and sound power are all based on spatially averaged values, some effects may be averaged out. Furthermore, measurement results are often rounded to the nearest tenth or hundredth value.[19,20] Other errors exist from theoretical assumptions alone. The reproducibility of reverberation chamber measurements also continues to be investigated.[20] The combined uncertainties from several factors in these measurements are likely to have a greater effect than the uncertainties present in the ρ_0 and c values alone.

V. SUMMARY AND CONCLUSIONS

This work has demonstrated that total acoustic energy density may be beneficially used in reverberation chamber measurements. The principal characteristic that makes it more attractive than squared pressure or potential energy density is its greater spatial uniformity over an enclosed sound field. Reverberation chamber measurements suffer from uncertainties associated with diffuse field approximations. Total energy density was expected to reduce measurement error or simplify measurement processes by requiring fewer measurement positions.

The work has specifically introduced methods of measuring sound absorption and sound power using total energy density. An energy density impulse response can be obtained by combining the squared-pressure impulse response and a squared-velocity-magnitude impulse response with appropriate weightings. The resulting response can then be used to determine reverberation time and sound absorption. The sound power level of a source can also be calculated in a reverberation chamber using the total sound energy density level (SEDL). Several procedures for implementing these measurements and the implications of atmospheric variable errors have been described in the paper.

Results from T_{60} measurements in two reverberation chambers show greater spatial uniformity of decay using the total energy density impulse response. In a large qualified chamber, the variation was roughly one-half that of the squared-pressure impulse response. The results suggest that fewer source-receiver positions are necessary to obtain an adequate sampling of the sound field. They also suggest that low-frequency T_{60} measurements are more consistent when using the energy density method. Similar results were found in a small reverberation chamber. The benefits automatically extend to the calculation of sound absorption values. Results from both chambers showed similar values for measured absorption coefficients that were comparable to those obtained from the standard squared-pressure method in the qualified chamber.

The sound power level of a reference sound source was also measured using both total energy density and squared pressure. Under steady-state conditions, total energy density was shown to have greater spatial uniformity than squared pressure in both chambers. With squared pressure, one source position, and six receiver positions, the large chamber failed to meet ISO requirements for variation in the 100-Hz third-octave band. This was in part because it was not fitted with low-frequency absorption. Nevertheless, it did meet the requirements when using total energy density and the same measurement positions. The variation of the latter was significantly lower than the maximum allowable variation at all frequencies of interest. This result was encouraging because it suggested that the low-frequency absorption and more source and receiver positions would not necessarily be required. It could also limit the need to reconfigure the chamber for different types of measurements. Results in the small chamber showed that the standard deviation values obtained from total energy density at low frequencies also met the ISO requirements, despite the fact that the room had a much smaller volume than required. Inclusion of the factors ρ_0 and c adds some uncertainty to SEDL measurements when atmospheric variables are not completely certain.

The investigation also revealed that the squared vector magnitude of the particle velocity often resulted in much smaller variations than squared pressure for both T_{60} values and sound levels. The degree of variation was quite similar to that of total energy density, especially at higher frequencies. In some cases, it was slightly better than that of total energy density. However, tests using both experimental and numerical data revealed that this observation was statistically insignificant. While the particle velocity measurements suggested that kinetic energy density alone could provide viable improvements to reverberation chamber measurements, it was found to be less consistent than total energy density over full measurement bandwidths.

This work has thus shown that total energy density can be used to successfully measure sound absorption and sound power in reverberation chambers with fewer measurement positions than required for squared pressure. It has also demonstrated that a fixed number of total energy density measurements may extend the usable low-frequency ranges of chambers with different volumes and limited low-frequency absorption (although this does not guarantee assumed diffuseness of their fields).

The investigation has suggested several topics that merit further investigation. For example, the study used a small number of discrete measurement positions according to ISO 354 and 3741. A thorough investigation of the sound field in a small enclosure would be beneficial to determine whether more accurate low-frequency measurements can be obtained in such a space by sampling the total energy density field with many more measurement positions than typically required in the standards. The impact of diffusers on the energy density methods should also be explored. The study was limited to sound absorption and sound power measurements.

The benefits of kinetic and total energy densities should also be explored for other types of reverberation chamber measurements, including sound transmission and sound scattering measurements. The authors encourage additional research in these areas.

ACKNOWLEDGMENTS

[1] J. Tichy and P. Baade, "Effect of rotating diffusers and sampling techniques on sound-pressure averaging in reverberation rooms," J. Acoust. Soc. Am. **56**, 137–143 (1974).

[2] R. K. Cook and P. A. Schade, "New method for measurement of the total energy density of sound waves," Proceedings of Inter-Noise 74, Washington DC, 1974, pp. 101–106.

[3] L. W. Sepmeyer and B. E. Walker, "Progress report on measurement of acoustic energy density in enclosed spaces," J. Acoust. Soc. Am. **55**, S12(A) (1974).

[4] R. V. Waterhouse and R. K. Cook, "Diffuse sound fields: Eigenmode and free-wave models," J. Acoust. Soc. Am. **59**, 576–581 (1976).

[5] R. V. Waterhouse, "Interference patterns in reverberant sound fields," J. Acoust. Soc. Am. **27**, 247–258 (1955).

[6] R. V. Waterhouse and R. K. Cook, "Interference patterns in reverberant sound fields. II," J. Acoust. Soc. Am. **37**, 424–428 (1965).

[7] F. Jacobsen, "The diffuse sound field: Statistical considerations concerning the reverberant field in the steady state," The Acoustics Laboratory, Technical University of Denmark, Report No. 27 (1979).

[8] J. A. Moryl, "A study of acoustic energy density in a reverberation room," M.S. thesis, The University of Texas at Austin, Austin, TX, 1987.

[9] J. A. Moryl and E. L. Hixson, "A total acoustic energy density sensor with applications to energy density measurement in a reverberation room," Proceedings of Inter-Noise 87, Beijing, China, 1987, Vol. II, pp. 1195–1198.

[10] J. Parkins, S. Sommerfeldt, and J. Tichy, "Narrowband and broadband active control in an enclosure using the acoustic energy density," J. Acoust. Soc. Am. **108**, 192–203 (2000).

[11] D. Bonsi, D. Gonzalez, and D. Stanzial, "Quadraphonic impulse responses for acoustic enhancement of audio tracks: measurement and analysis," Forum Acusticum (2005), pp. 335–340.

[12] I. Wolff and F. Massa, "Use of pressure gradient microphones for acoustical measurements," J. Acoust. Soc. Am. **4**, 217–234 (1933).

[13] F. Fahy, "Measurement of acoustic intensity using the cross-spectral density of two microphone signals," J. Acoust. Soc. Am. **62**, 1057–1059 (1977).

[14] F. J. Fahy, *Sound Intensity*, 2nd ed. (E & FN Spon, London, 1995).

[15] J. Ghan, "Expression for the estimation of time-averaged acoustic energy density using the two-microphone method (L)," J. Acoust. Soc. Am. **113**, 2404–2407 (2003).

[16] G. W. Elko, "An acoustic vector-field probe with calculable obstacle bias," Proc. Noise-Con **91**(7), 525–532 (1991).

[17] H-E. de Bree, "An overview of Microflown technologies," Acta Acust. **31**, 91–94 (2003).

[18] R. Raangs, Exploring the use of the Microflown, Ph.D. dissertation, University of Twente, 2005.

[19] ISO 3741:1999(E), "Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for reverberation rooms" (International Organization for Standardization, Geneva, 1999).

[20] ISO 354:2003, "Acoustics—Measurement of sound absorption in a reverberation room" (International Organization for Standardization, Geneva, 2003).

[21] A. Schaffner, "Accurate estimation of the mean sound pressure level in enclosures," J. Acoust. Soc. Am. **106**(2), 823–827 (1999).

[22] M. Vorländer, "Revised relation between the sound power and the average sound pressure level in rooms and the consequences for acoustic measurements," Acustica **81**, 332–343 (1995).

[23] R. V. Waterhouse, "Statistical properties of reverberant sound fields," J. Acoust. Soc. Am. **43**, 1436–1444 (1968).

[24] D. Lubman, "Fluctuations of sound with position in a reverberant room," J. Acoust. Soc. Am. **44**, 1491–1502 (1968).

[25] J. L. Davy, I. P. Dunn, and P. Dubout, "The Variance of Decay Rates in Reverberation Rooms," Acustica **43**, 12–25 (1979).

[26] C. Eyring, "Reverberation time in dead rooms," J. Acoust. Soc. Am. **1**, 217–214 (1930).

[27] M. Hodgson, "On measures to increase sound-field diffuseness and the applicability of diffuse-field theory," J. Acoust. Soc. Am. **95**(6), 3651–3654 (1994).

[28] K. H. Kutruff, "Sound decay in reverberation chambers with diffusing elements," J. Acoust. Soc. Am. **69**, 1716–1723 (1981).

[29] M. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).

[30] D. Nutter, "Sound absorption and sound power measurements in reverberation chambers using energy density methods," M.S. thesis, Brigham Young University, 2006.

[31] ISO 9613-1:1993(E), "Acoustics—Attenuation of sound during propagation outdoors—Part 1: Calculation of the absorption of sound by the atmosphere" (International Organization for Standardization, Geneva, 1993).

[32] A. Pierce, *Acoustics, An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Melville, NY, 1989).

[33] G. Mange, "The effect of mean free path on reverberation room measurement of absorption and absorption coefficients," Noise Control Eng. J. **53**(6), 268–270 (2005).

[34] ANSI S1.1-1994, "Acoustical terminology" (American National Standards Institute, New York, 1994).

[35] D. Stanzial, D. Bonsi, and N. Prodi, "Measurement of new energetic parameters for the objective characterization of an opera house," J. Sound Vib. **232**, 192–211 (2000).

[36] G. Keller, *Applied Statistics with Microsoft® Excel* (Duxbury, Pacific Grove, CA, 2001).

[37] H. Kuttruff, *Room Acoustics*, 4th ed. (Spon, New York, 2000).

[38] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).

Analysis of the performance of a particle velocity sensor between two cylindrical obstructions

J. W. van Honschoten,[a] D. R. Yntema, V. B. Svetovoy, M. Dijkstra,
R. J. Wiegerink, and M. Elwenspoek
*Transducers Science and Technology Group, MESA$^+$ Institute for Nanotechnology, University of Twente,
The Netherlands*

The performance of an acoustic particle velocity sensor that is placed between two cylindrical objects has been analyzed both analytically and by means of finite volume simulations on fluid dynamics. The results are compared with acoustic experiments that show a large magnification of the output signal of the particle velocity sensor due to the mounting of the sensor between two cylinders. The influences of this construction consist of an attenuation of particle velocities at frequencies below a few hertz, whereas signals in the higher frequency range are amplified, up to approximately three times (10 dB) in a frequency range between 50 and 1000 Hz. The theoretical analysis is based on the derivation of the stream function for the situation of two long cylinders immersed in an oscillating incompressible viscous fluid, at low Reynolds numbers. The results lead to an improved insight into the effects of viscosity and fluid flow that play a role in acoustic measurements and open the way for further optimization of the sensitivity of the sensor. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717406]

I. INTRODUCTION

This paper discusses a micromachined acoustic sensor based on a thermal measurement principle that measures particle velocity instead of sound pressure, the scalar quantity that is measured by conventional microphones.[1,2] The sensor has been optimized for sound measurement purposes, and it has been applied to the measurement of one- and three-dimensional sound intensities,[2] acoustic impedances,[3,4] and far field pressure.[5] Other applications of the sensor such as an add-on microphone for professional recording purposes have also been shown.[6] An important property of the particle velocity sensor, contrary to pressure gradient microphones, is its comparatively high sensitivity to low frequency sound waves; it can thus be used as well for the measurement of dc flows,[7] and it can be applied as a mass flow sensor.

A thermal particle velocity sensor usually consists of two closely spaced (spacing about 100 μm) thin wires of silicon nitride, with an electrically conducting platinum pattern on top of them, as shown in Fig. 1. The dimensions of the two wires are $1500 \times 2 \times 0.3$ μm. The platinum layer on top of the silicon nitride is used as a heater *and* as a temperature sensor by using the temperature-dependent resistance of the platinum. The wires are heated by an electrical current to an operation temperature between 300 and 400 °C. The particle velocity associated with the sound wave will modulate the temperature distribution around the resistors, which can be measured electronically due to the related change in resistance values.

In 2000, a theoretical description and a model for the behavior of the device were presented[8] and an explicit analytical expression for the sensitivity was deduced.[9,10]

For measurement purposes, the particle velocity sensor is placed in a protective package: an approximately 7-cm-long cylindrical probe of 13 mm diameter with two small cylinders of 5 mm diameter at its end, with the sensor in between [see Fig. 2 for a picture of the commercially available "PU" probe and Figs. 3(b) and 3(c) for a model representation]. This packaging of the sensor improves its sensitivity considerably. In the case of our experimental device the sensitivity was increased by a factor of 3.2, or approximately 10 dB at 500 Hz. This observation raised the need for a detailed investigation of the effects of the two adjoining cylindrical objects of the package, in order to optimize the sensor performance further. In this paper, this investigation is achieved both by a theoretical description of the flow profile around two long cylinders and by a numerical analysis by means of finite-volume computational fluid dynamic simulations on this two cylinder geometry and on the specific probe package with its two small cylindrical tubes.

The phenomenon of oscillatory viscous flow around a long cylinder has been subject of investigation for many years, especially for high Reynolds numbers. One of the points of interest has been the time-independent streaming motion as a result of the interaction of the oscillatory viscous flow with a solid boundary. This induced steady streaming is generated by the nonlinear Reynold stresses in the unsteady boundary layer. For Reynolds numbers Re ∼ 10 the steady streaming near an oscillating circular cylinder was studied by Carrière,[11] while Andrade[12] and Schlichting[13] did experiments in the range Re ∼ 1000. A theory in terms of expansions in the inner and outer regions around the cylinder was proposed by Riley.[14] Wang[15] solved the problem of oscillatory viscous flows around a single cylinder, for a broad range of Reynolds and Strouhal numbers. Zapranov *et al.*[16] ana-

---

[a] Electronic mail: j.w.vanhonschoten@utwente.nl
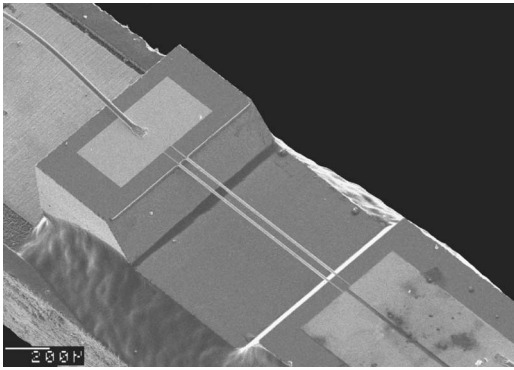
FIG. 1. SEM photograph of a two-wire type particle velocity sensor.

lyzed the flow around two parallel cylinders with the plane of their axes both perpendicular and parallel to the flow, whereas they concentrated on the high Reynolds number range when steady streaming becomes important. Flow around two parallel cylinders placed one behind the other in the propagation of an acoustic sound wave was studied by Zhuk et al.[17] Two parallel cylinders in an oscillatory flow have also been investigated numerically,[18] but especially for relatively high Reynolds numbers (Re~200).

In this paper, however, we concentrate on the oscillatory flow due to a sound wave around two parallel cylinders in the range of low Reynolds numbers (Re~1) since the typical velocities associated with sound waves are relatively small. The equations to be solved then become linear.[19] We follow broadly the approach proposed by Zapryanov et al.,[16] but investigate especially the fluid flow for small particle

FIG. 2. The commercially available PU probe, with the particle velocity sensor packaged in between the two cylinders.
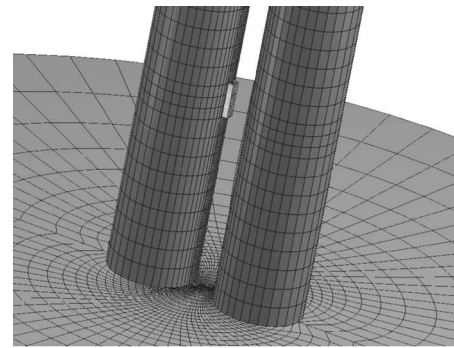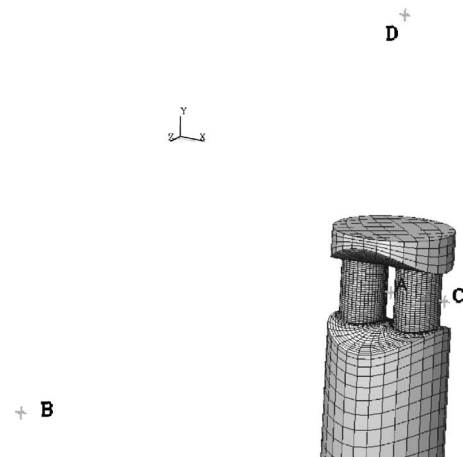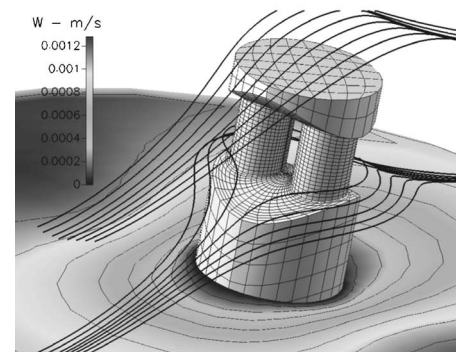
FIG. 3. (a) The solution space for numerical calculations on two long cylinders with a small rectangular obstacle at fixed temperature in between. The geometry consists of 50 000 cells. (b) A grid containing about 70 000 cells is defined to model the probe geometry, very dense structured at the place of interest (the center). Four points in space were defined in particular: point A: in between the two cylinders, B: at large distance in front of the probe, C: left from the probe, 6.5 mm from the center, D: at 3 cm above A. (c) Simulation result visualizing both the instantaneous streamlines (defined at a particular time $t=t_0$) and a contour plot of the particle velocity at $v_0 = 1$ mm/s; $f=1$ Hz.

velocities. The assumption that we shall make is that the characteristic dimensions of the problem (~ 15 mm) are much smaller than the acoustic wavelength, which is valid for frequencies much lower than 20 kHz. In this case the propagation of interactions in the gas can be regarded as instantaneous, which, together with the condition that the fluid velocity be small compared with that of sound, implies that the fluid can be described as incompressible.

van Honschoten et al.: Particle velocity sensor between cylindrical obstructions

II. THEORY

A. Introduction

For a description of the flow behavior around the sensor, the full Navier-Stokes equations for the three-dimensional geometry should be solved. For the complex geometry of the actual probe as shown in Fig. 2, an exact solution cannot be found.

However, the two cylindrical obstructions of this package on both sides of the particle velocity sensor raise the need for an analysis of the acoustic flow profile around two parallel long cylinders, a problem that is actually two dimensional. As will be shown in the following, the solution to this two-dimensional problem offers a quite adequate description of the more complicated probe geometry.

B. Assumptions and problem definition

To find the flow profile in and around the geometry of interest, we have to solve the equations of motion of a viscous fluid, the Navier-Stokes equations. In their most general form, these equations are rarely solvable for complicated geometries. Therefore, several assumptions to simplify the problem and some approximations have to be made.

In the following approach the influence of the sensor between the two cylinders on the flow behavior will be neglected, and the fluid is considered as incompressible. In that case the problem consists of a viscous fluid of constant density ρ and kinematic viscosity ν in which two separated parallel cylinders are immersed. At infinity, the fluid oscillates harmonically, perpendicular to the plane containing their axes, with a velocity $u_0 \cos \omega t$. Two regions of interest are distinguished: a frequency range $\omega \ll \omega_c$, in which viscous effects will be seen to be dominant and the viscous boundary layers around the cylinders become large, and a region $\omega \gg \omega_c$, where the fluid behavior approaches that of an ideal gas.

To determine if the fluid in the case of propagating acoustic waves can be regarded as incompressible, we consider the conditions under which the assumption of incompressibility is justified. The fluid may be regarded as incompressible if $\Delta\rho/\rho \ll 1$. A necessary condition for this is that the fluid velocity be small compared with that of sound:[20]

$$v \ll c. \tag{1}$$

However, this condition is sufficient only in steady flow. In nonsteady flow, a further condition must be fulfilled. If τ is a characteristic time over which the fluid velocity undergoes significant changes, then the second condition reads

$$\tau \gg l/c. \tag{2}$$

Condition (2) means in fact that the time l/c taken by a sound signal to traverse the distance l must be small compared with the time τ during which the flow changes appreciably, so that the propagation of interactions in the gas may be regarded as instantaneous. The characteristic length l of the problem is orders of magnitude smaller than the acoustic wave length.

Typical values for the magnitude of the particle velocity of sound waves are $v = 10^{-5} - 10^{-2}$ m s^{-1}. Taking $v = 2.5$

$\times 10^{-3}$ m s^{-1} (~94 dB sound level), $c = 350$ m s^{-1}, $l = 5$ mm, we see that for all frequencies well below 10 kHz, conditions (1) and (2) are fulfilled, and one can describe the gas as incompressible.

In their most general form the Navier-Stokes equations for an incompressible fluid then read

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \boldsymbol{\nabla})\mathbf{v} = -\frac{1}{\rho} \boldsymbol{\nabla} p + \nu \nabla^2 \mathbf{v} \tag{3}$$

with $\mathbf{v}$ the (vectorial) velocity, p the pressure, ν the kinematic viscosity, and ρ the fluid density.

Besides, the continuity equation has to be obeyed:

$$\frac{\partial \rho}{\partial t} + \rho \boldsymbol{\nabla} \cdot \mathbf{v} = 0. \tag{4}$$

Let us assume an infinite incompressible viscous fluid in which two parallel circular cylinders of radius R are immersed. The presence of the sensor itself and its high temperature are neglected as yet, later we will see the justification for this. The fluid oscillates in the direction perpendicular to the plane containing the axes of the cylinders with velocity $u_0 \cos 2\pi f t$ at infinity where u_0 is the magnitude of the particle velocity and f the frequency of the wave.

The problem is scaled using the dimensionless parameters

$$\tau = \frac{\nu}{l^2} t, \quad \mu = \frac{2\pi f l^2}{\nu}, \tag{5}$$

where l represents a characteristic length, e.g., the cylinder radius.

The Navier-Stokes equations in the form (3) are nonlinear because of the second term on the left-hand side. This nonlinear problem has been analyzed by Zapryanov et al.[16] They used a perturbation theory in terms of asymptotic expansions in the inner and outer regions around the cylinders.[15,16]

We can estimate the nonlinear convection term $(\mathbf{v} \cdot \boldsymbol{\nabla})\mathbf{v}$ to be in the order of u_0^2/l, while the magnitude of the time-dependent term can be approximated as ωu_0 and the viscous term as $\nu u_0/l^2$. One sees that the nonlinear term can be neglected compared with the other terms in the Navier-Stokes equations if

$$u_0 \ll \max\left\{\frac{\nu}{l}, 2\pi f l\right\}. \tag{6}$$

For the current values in our problem, $u_0 \sim 2 \times 10^{-4}$ m/s, $l \sim 6 \times 10^{-3}$ m, and $\nu \sim 1.5 \times 10^{-5}$ m^2/s, corresponding to a Reynolds number of Re $\ll 1$. Taking as a lower limit for the frequency 1 Hz, we see that $\nu/l \approx 3 \times 10^{-3}$ m/s and $2\pi f l \approx 4 \times 10^{-2}$ m/s so that condition (6) is well fulfilled.

Further, we assume $\mu \gg 1$.

Taking the curl of both sides of Eq. (3) and using the equation $\boldsymbol{\nabla} \cdot \mathbf{v} = 0$, we get
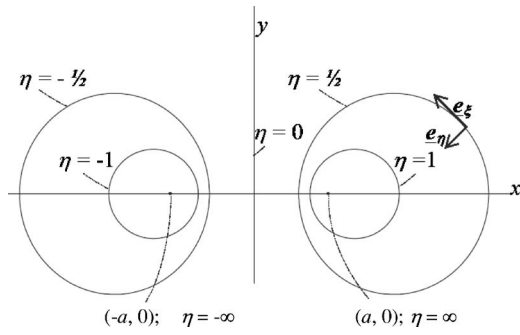
FIG. 4. Representation of the used coordinate system.

$$\frac{\partial}{\partial \tau}(\nabla \times \mathbf{v}) + (\mathbf{v} \cdot \nabla)\, \nabla \times \mathbf{v} - (\nabla \times \mathbf{v} \cdot \nabla)v = \frac{1}{\mu}\Delta(\nabla \times \mathbf{v}). \tag{7}$$

The equations of motion, Eq. (3), are often written in terms of the stream function Ψ,[16,20] defined by

$$\nu_x = \frac{\partial \Psi}{\partial y}, \quad \nu_y = -\frac{\partial \Psi}{\partial x}. \tag{8}$$

Neglecting now the nonlinear terms, one obtains with Eqs. (7) and (8) the equation to be solved for the stream function

$$\frac{\partial}{\partial \tau}(\Delta \Psi) - \frac{1}{\mu}\Delta^2 \Psi = 0. \tag{9}$$

C. Solution

The problem can be well described in a two-dimensional system of bipolar cylindrical coordinates (ξ, η), defined by the transformations

$$x = \frac{a \sinh \eta}{\cosh \eta - \cos \xi}, \quad y = \frac{a \sin \xi}{\cosh \eta - \cos \xi}, \quad a = 1, \tag{10}$$

where x and y are the usual Cartesian coordinates, $0 \leq \xi < 2\pi$ and $-\infty < \eta < \infty$. See Fig. 4. The value of the length parameter a is in principle arbitrary, for convenience we have chosen it as yet to be $a = 1$. The two cylinders are therefore defined by $\eta = \eta_1 > 0$ and $\eta = \eta_2 < 0$, such that $R = a/|\sinh \eta_1| = a/|\sinh \eta_2|$ and we consider the symmetric case $\eta_1 = -\eta_2$. The fluid region is given by $\eta_2 < \eta < \eta_1$, $0 \leq \xi < 2\pi$, while $\eta = \xi = 0$ at infinity.

In this coordinate system, the stream function Ψ is defined by

$$\nu_\xi = (\cosh \eta - \cos \xi)\frac{\partial \Psi}{\partial \eta}, \quad \nu_\eta = -(\cosh \eta - \cos \xi)\frac{\partial \Psi}{\partial \xi}. \tag{11}$$

The equation for the stream function,

$$\frac{\partial}{\partial \tau}(\Delta \Psi) - \frac{1}{\mu}\Delta^2 \Psi = 0, \tag{12}$$

is to be solved in the (ξ, η) plane, with Δ the Laplacian operator in this coordinate system, i.e.,

$$\Delta = (\cosh \eta - \cos \xi)^2 \left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} \right). \tag{13}$$

If $\psi_1 - \psi_2$ is the (dimensionless) total flux between the two cylinders, the boundary conditions for Ψ become

$$\Psi = \psi_i, \quad \frac{\partial \Psi}{\partial \eta} = 0 \quad \text{at } \eta = \eta_i \quad (i = 1, 2), \tag{14}$$

$$\Psi = \frac{-\sinh \eta}{\cosh \eta - \cos \xi}e^{i\tau} \quad \text{at } \eta \to 0, \quad \xi \to 0. \tag{15}$$

We assume an oscillatory flow with time dependence $e^{i\tau}, \tau = 2\pi ft$, and eliminate this time dependence in the equations by defining $\Psi = \bar{\Psi}(\xi, \eta, \tau)e^{i\tau}$. This results in

$$\frac{\partial}{\partial \tau}(\Delta \bar{\Psi}) + i\Delta \bar{\Psi} - \frac{1}{\mu}\Delta^2 \bar{\Psi} = 0. \tag{16}$$

For convenience, the bar on Ψ is omitted from now on, and after neglecting the slow time dependence, the equation to be solved becomes

$$\Delta(\Delta \Psi - i\mu \Psi) = 0. \tag{17}$$

To do this, we divide the fluid region into separate regions: two thin boundary layers around both cylinders, and an intermediate region outside these layers. Besides, the problem is symmetric in the axis $\eta = 0$ (we consider the case $\eta_1 = \eta_2$), so we restrict ourselves to the region $\eta > 0$ (with only one cylinder and one boundary layer) and realize that the final stream function is an antisymmetric function in η.

In the region outside the boundary layer, we can neglect viscous effects and make use of the assumption that μ is large, so that the equation for the stream function to be solved reduces to the Laplace equation:

$$\Delta \Psi = 0 \tag{18}$$

including the boundary conditions (14) and (15).

With the following coordinate transformation to p and q:

$$p = \frac{\xi + i\eta}{2}, \quad q = \frac{\xi - i\eta}{2} \tag{19}$$

so that $(\cosh \eta - \cos \xi) = 2 \sin p \sin q$, the Laplace equation in p, q coordinates reads

$$\sin^2 p \sin^2 q \frac{\partial^2}{\partial p \partial q}\Psi = 0 \tag{20}$$

we see that every function of the form $\Psi(p, q) = f_1(p) + f_2(q)$ is a solution, where f_1 and f_2 are functions that depend only on p and only on q, respectively. Since the boundary condition (15) can be written as $\Psi = -(i/2)(\cot p - \cot q)$ for $p \to 0, q \to 0$, the stream function

$$\Psi = -\frac{i}{2}\left(\cot \frac{\xi + i\eta}{2} - \cot \frac{\xi - i\eta}{2} \right) = \frac{-\sinh \eta}{\cosh \eta - \cos \xi} \tag{21}$$

obeys the Laplace equation, Eq. (18), and condition (15). To meet the other boundary condition, Eq. (14), we add two functions $f_1(p)$ and $f_2(q)$, that have to be determined from the requirement:

van Honschoten *et al.*: Particle velocity sensor between cylindrical obstructions

$$\Psi = \Psi_0 \quad \text{at } \eta = \eta_1 \tag{22}$$

with Ψ_0 a constant. Besides, Ψ is 2π periodic in ξ.

Writing f_1 and f_2 in the form of expansions,

$$f_1\left(\frac{\xi + i\eta}{2}\right) = \sum_{n=-\infty}^{\infty} a_n \exp\left(2in\frac{\xi + i\eta}{2}\right),$$

$$f_2\left(\frac{\xi - i\eta}{2}\right) = \sum_{n=-\infty}^{\infty} b_n \exp\left(2in\frac{\xi - i\eta}{2}\right), \tag{23}$$

the sum of them, $f_1((\xi+i\eta)/2)+f_2((\xi-i\eta)/2)$, can be written in the form

$$\sum_{n=-\infty}^{\infty} (a_n e^{-n\eta} + b_n e^{n\eta}) e^{in\xi}. \tag{24}$$

This infinite series should converge for all $0 \leq \eta \leq \eta_1$, $0 \leq \xi \leq 2\pi$ and moreover, the final function has to be even in ξ. These requirements lead to the following expression for the series:

$$f_1\left(\frac{\xi + i\eta}{2}\right) + f_2\left(\frac{\xi - i\eta}{2}\right) = A_0 + \sum_{n=0}^{\infty} A_n e^{n\eta} \cos n\xi, \tag{25}$$

where the coefficients A_n follow from the boundary conditions at $\eta = \eta_1$:

$$A_n = \frac{\exp(-n\eta_1)}{\pi} \int_{-\pi}^{\pi} \left(\frac{\sinh \eta_1}{\cosh \eta_1 - \cos \xi}\right) \cos n\xi d\xi \tag{26}$$

The coefficients A_n, as expressed by Eq. (26) can be calculated explicitly[21] and become finally very simple:

$$A_n = 2 \exp(-2n\eta_1). \tag{27}$$

We thus find for the total stream function (that is defined up to an arbitrary constant) in the region outside the boundary layers:

$$\Psi(\xi, \eta) = \left(-\frac{\sinh \eta}{\cosh \eta - \cos \xi} + \sum_{n=1}^{\infty} A_n \exp(n\eta)\cos n\xi\right) \exp(i\tau) \tag{28}$$

with A_n as defined in Eq. (27).

With this expression for A_n the sum in Eq. (28) can be calculated straightforwardly and the stream function in the region outside the boundary layers becomes

$$\Psi(\xi, \eta) = \left(\frac{\exp(\eta - 2\eta_1 + i\xi)}{1 - \exp(\eta - 2\eta_1 + i\xi)} + \frac{\exp(\eta - 2\eta_1 - i\xi)}{1 - \exp(\eta - 2\eta_1 - i\xi)} - \frac{\sinh \eta}{\cosh \eta - \cos \xi}\right) \exp(i\tau). \tag{29}$$

Using the inverse transformation of Eq. (10), this can be written in Cartesian coordinates as

$$\Psi(x,y) = 2 \operatorname{Re}\left(\left(\frac{y + i(x + a)}{y(\exp(2\eta_1) - 1) + i(x(\exp(2\eta_1 - 1)) - a(\exp(2\eta_1) + 1))}\right) - x\right) \exp(i\tau). \tag{30}$$

In the boundary layer adjacent to the right cylinder, Eq. (17) is solved by finding the solution of

$$\Delta\Psi_b - i\mu\Psi_b = 0 \tag{31}$$

with Ψ_b the stream function in this region and μ large but finite.

One can now introduce the scaled variable ζ in this region:

$$\zeta = (\eta_1 - \eta)\sqrt{\mu}. \tag{32}$$

The boundary layer extends from $\zeta=0$ to $\zeta=1$.

Since in this boundary layer η does not change significantly, and because the ξ derivatives of Ψ are small compared with the η derivatives, the equation to be solved becomes

$$\frac{\partial^2 \Psi_b}{\partial \zeta^2} - i(\cosh \eta_1 - \cos \xi)^2 \frac{\partial^4 \Psi_b}{\partial \zeta^4} = 0, \tag{33}$$

which has the solution

$$\Psi_b(\xi, \zeta) = C_1 \exp(\sqrt{i}(\cosh \eta_1 - \cos \xi)\zeta) + C_2 \exp(-\sqrt{i}(\cosh \eta_1 - \cos \xi)\zeta) + C_3\zeta + C_4. \tag{34}$$

The coefficients C_1, C_2, and C_3 are to be determined from the requirements for Ψ_b that

$$\Psi_b = \psi_1, \quad \frac{\partial \Psi_b}{\partial \zeta} = 0 \quad \text{at } \zeta = 0 \quad (\eta = \eta_1), \tag{35}$$

$$\frac{\partial \Psi_b}{\partial \zeta}\frac{d\zeta}{d\eta} = \frac{\partial \Psi}{\partial \eta} \quad \text{at } \zeta = 1 \quad (\eta = \eta_1 + 1/\sqrt{\mu}) \tag{36}$$

and, since the velocity v_ξ should also be continuous,

van Honschoten *et al.*: Particle velocity sensor between cylindrical obstructions 2715
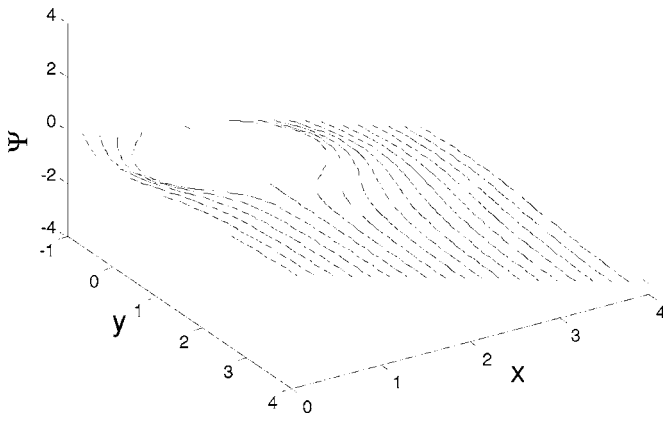
FIG. 5. Stream function Ψ as calculated from Eq. (29). Contour lines of Ψ form the streamlines in the flow.

FIG. 6. Velocity v_y (normalized to u_0) along the line $y=0$, for the two cylinder geometry with $R=2.47$ mm and $d=1.56$ mm, at $f=2.5$ Hz and $f=4$ Hz, according to Eqs. (29) and (34).

$$M^2 \frac{\partial^2 \Psi_b}{\partial \zeta^2} = \frac{\partial^2 \Psi}{\partial \eta^2} \quad \text{at } \zeta = 1 \quad (\eta = \eta_1 + 1/\sqrt{\mu}). \quad (37)$$

Explicit expressions for the coefficients $C_1(\xi, \eta_1)$, $C_2(\xi, \eta_1)$, and $C_3(\xi, \eta_1)$ follow from a straightforward calculation from Eqs. (35), (36), and (37) but are not shown here for brevity. For a specific value of η_1 the result of Eq. (34) with these values of the coefficients is plotted in Fig. 6.

D. Discussion

The fluid region around the two cylinders has been divided in three regions: The two boundary layers adjacent to the cylinders, of thickness $\delta \sim R\mu^{-1/2}$, and a region in between. The Navier-Stokes equations are solved in terms of a stream function Ψ. In the boundary regions the stream function Ψ_b is governed by the parameter μ, and this solution is matched to the function Ψ in the intermediate region.

In Fig. 5, the stream function according to Eq. (29) is visualized in the x-y plane, for the half-infinite region $x>0$ and a cylinder radius $R=\sinh \eta_1$ ($a=\eta_1=1$). Contour lines of Ψ form the streamlines of the flow.

Since the reduced frequency parameter $\mu = \pi f l^2/\nu$ is proportional to frequency f, the situation $\mu \to \infty$ describes the nonviscous fluid limit, and the equation to be solved becomes the Laplace equation, Eq. (18). The ideal case of one infinitely long cylinder of radius R in an incompressible nonviscous ideal fluid with a velocity $u_0 e^{i2\pi ft}$ at infinity is relatively easily solved in radial coordinates, see, e.g., Landau.[20] As is also the case in our situation, this problem is two dimensional and it can be solved in terms of a potential function Φ and stream function Ψ (Φ and Ψ are the real and imaginary parts of the complex potential w, respectively). One then finds for the potential Φ and the velocity $\mathbf{v}$ of one cylinder immersed in a nonviscous fluid

$$\Phi = \frac{R^2}{r}\mathbf{u}_0 \cdot \mathbf{n} + \mathbf{u}_0 \cdot \mathbf{r}, \quad \mathbf{v} = \mathbf{u} - \frac{R^2}{r^2}(2\mathbf{n}(\mathbf{u}_0 \cdot \mathbf{n}) - \mathbf{u}_0) \quad (38)$$

with $\mathbf{n}$ a unit vector in the direction of $\mathbf{r}$ and with the origin as the center of the cylinder. We recognize in this expression the contribution of the uniform flow (the term $\mathbf{u}_0 \cdot \mathbf{r}$ in the expression for Φ), and a "doublet flow"[22] or "dipole" solution (the term $\mathbf{u}_0 \cdot \mathbf{n}/r$).
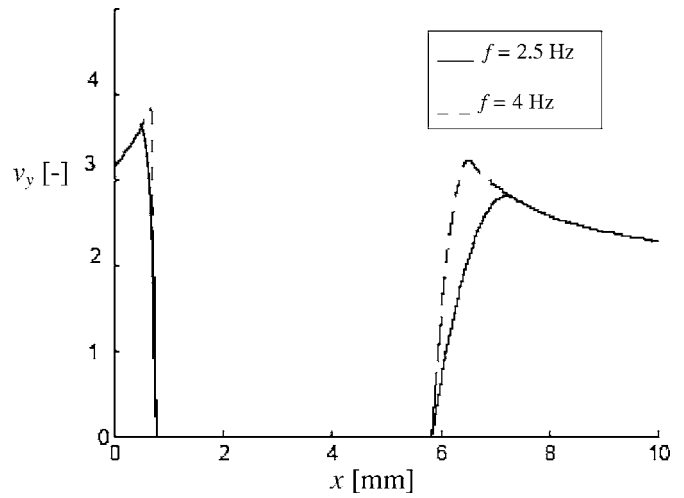
In the two-dimensional problem of fluid flow around two circle-cylinders, the situation becomes more complicated, but the fluid flow can still be approximated by a composition of elementary plane flows. A suitable, equivalent, approach is to write a proper superposition of uniform and dipole solutions. As mentioned, the potential flow about a single cylinder in a uniform flow field is composed of the basic expressions for a parallel flow and for a dipole of given strength. To describe the nonviscous flow around multiple circular cylinders, the so-called method of images of Milne-Thomson for circle-cylinders can be utilized. With this method the boundary condition on the surface of the circle-cylinders can be satisfied to an arbitrary order of approximation.[23] Returning now to our solution, we can also recognize in Eq. (30) these components. Obviously, the term—x represents the uniform flow. Investigating the second term of this expression in the limit for large $z=x+iy$, we can see that it behaves as a sum of dipole solutions.

In the other limit, when the frequency parameter $\mu \to 1$, viscous effects will dominate. The presented matching solutions approach remains valid as long as μ is large, so that the boundary layer $\delta \sim R\mu^{-1/2}$ is small compared to the other dimensions, the cylinder radius and the mutual cylinder distance d. If δ becomes of the order of d, the boundary layer extends over all spacing between the cylinders of the probe and the used method does not apply. In that case, the two cylinders reduce to one object for acoustic waves imposed on the probe.

E. Consequences of the model

From the obtained expression for the stream function in the boundary layer and in the intermediate region, Eqs. (29) and (34), the velocity ν_y can be calculated. We are particularly interested in its value along the line $y=0$.

This velocity, both in the intermediate region and in the thin boundary layers, is depicted in Fig. 6. It was calculated for the two cylinder geometry with a radius $R_0=2.47$ mm and a spacing of $d_0=1.56$ mm, and two values of the parameter μ, corresponding to $f=2.5$ Hz and $f=4$ Hz. For even

van Honschoten *et al.*: Particle velocity sensor between cylindrical obstructions
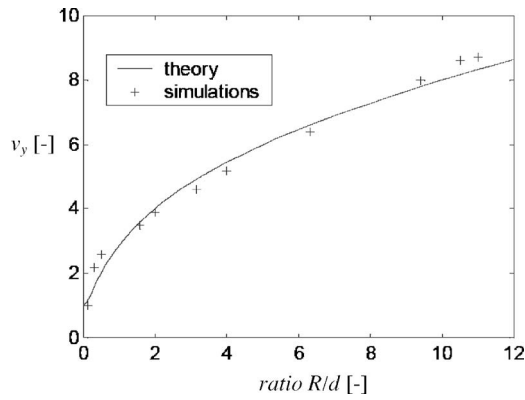
FIG. 7. Normalized velocity v_y at the central point $(x,y)=(0,0)$ as a function of the ratio R/d, according to theory and numerical simulations on the probe geometry with $\mu=10$.

lower frequencies, the boundary layer thickness is in the order of the mutual cylinder spacing and the approach leading to Eq. (34) is no longer valid. Let us now focus on the magnitude of the y velocity at the location of the particle velocity sensor, i.e., at the point $(x,y)=(0,0)$. In particular we will investigate the velocity in relation to the ratio of the cylinder radius and the spacing between the cylinders. With the coordinate transformation of Eq. (10) the mutual distance d between the cylinders, see Fig. 4, can be expressed as (with $a=1$):

$$d = \frac{2 \sinh \eta_1}{\cosh \eta_1 + 1} \qquad (39)$$

while the cylinder radius was $R=1/\sinh \eta_1$. The ratio R/d is therefore

$$\frac{R}{d} = \frac{1}{2(\cosh \eta_1 - 1)}. \qquad (40)$$

From Eq. (29) and applying Eq. (11) we can determine the velocity v_y at $(x,y)=(0,0)$, where $(\xi,\eta)=(\pi,0)$. In Fig. 7 the dependence of $v_y(0,0)$ on the ratio R/d according to Eq. (29) is plotted. Also shown are numerical results from simulations on the real geometry of the probe with varying R/d, as will be discussed in more detail in Sec. III, rescaled to a constant value of the normalized frequency parameter μ to take into account the boundary layer effects. The value of this parameter, defined as $\mu=2\pi fl^2/v$, where for the characteristic dimension l was chosen the mutual cylinder distance d, was kept at a constant value of $\mu=10$. It must be mentioned that the theoretical curve is only valid as long as the probe dimensions, d and R, remain small with respect to the acoustic wavelength: $R, d \ll \lambda$.

In Fig. 8 the velocity is depicted as a function of frequency. The graph shows the dependence of the velocity on the normalized frequency parameter μ (upper axis) as well, indicating that the theory yields a characteristic frequency of about $\mu_c=4$. For $\mu>\mu_c$, the magnification remains almost constant.

The fact that high frequencies are amplified by a factor that does not depend much on frequency in the frequency range of our main interest, 50–1000 Hz, and that low frequency signals are attenuated due to the viscous effects, is of
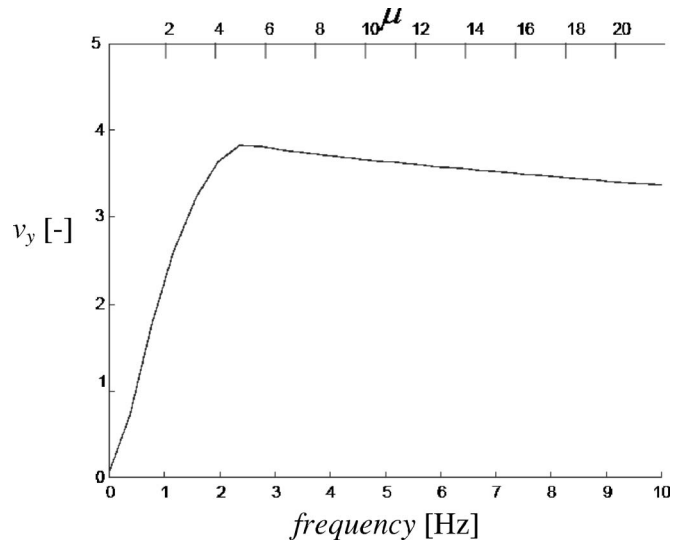
FIG. 8. Dependence of v_y on frequency, at the point where the sensor is located between the cylinders, $(x,y)=(500 \ \mu\text{m},0)$. Calculated according to Eqs. (29) and (34).

great importance for measurement applications. Low frequency signals, ubiquitous for this sensor,[24] should be suppressed while signals above a certain frequency are to be amplified. The package thus significantly contributes to the low-frequency roll off below about 10 Hz, which can be realized electronically also (actually most commercial probes have such an electronic high-pass filter).

III. COMPUTATIONAL FLUID DYNAMICS

A. The setup

To verify the validity of the theoretical analysis and the correctness of the used assumptions, we performed numerical simulations on two parallel cylinders immersed in a compressible flow, with the small particle velocity sensor in between, at a sensor temperature of 700 K.

For the numerical calculations on these structures we used CFDRC, a commercially available software program for fluid dynamical simulations.[25,26] The software provides a variety of tools for the simulation and analysis of fluid flow. In our approach for the numerical simulation of the fluid behavior around the sensor, three successive steps are to be distinguished. First, the volume of interest (the solution space) was divided into discrete control volumes or cells. Second, the boundary conditions, the initial conditions, and the equations to be solved at each cell were defined, as well as the numerical technique to solve the equations. And finally, after the solution, we extracted the needed information from the large amount of data generated in the solution process.

The solution space was defined as a system of two infinitely long cylinders of $R_0=2.47$ mm diameter at an ambient temperature of 300 K, with a small rectangular element of $0.2\times0.2\times2$ mm, representing a heated particle velocity sensor, of a fixed temperature of 700 K in between. The infiniteness of the cylinders was realized by imposing periodic boundary conditions on the flow. The fluid space around the probe was meshed using a structured grid of tetrahedral and

prismatic volume elements. The number of cells amounted to about 50 000; in the middle, around the sensor, the cells were made very dense. See Fig. 3(a).

Additionally, a solution space was defined with respect to the commercially available PU probe as shown in Fig. 2, the probe that was one of the motivations for this numerical analysis. This geometry consisted of a cylinder of approximately 8 cm radius and 15 cm height, in which the probe was positioned. This solution space comprised about 70 000 cells, with a dense grid in the middle around the sensor. See Fig. 3(b).

As a boundary condition, a plane propagating wave was imposed on the boundary of the large outer cylinder. This wave was described by a varying fluid particle velocity of magnitude u_0 and frequency f: $u(y,t)=u_0 \cos(2\pi ft - ky)$, with k the wave number in the propagation direction y. The Navier-Stokes equations were solved at each fluid space element, together with the no-slip boundary condition on the probe surface and the assumption of a fully adiabatic process. The calculations were performed with a convergence criterion of 10^{-4}, using the SIMPLEC solution method, coupled with the ideal gas law.[25]

Besides, a constant ambient temperature and constant dynamic viscosity of, respectively, T=300 K and ν =1.5895×10^{-5} m²/s were assumed, an equilibrium fluid density of ρ=1.1614 kg/m³, and an equilibrium pressure of p_0=1.0×10^5 Pa in the fluid around the probe. It must be mentioned that both the presence of the particle velocity sensor and the temperature effects of this sensor, that causes a local temperature increase of the fluid due to the heated wires, have not been taken into account in the simulations on the probe geometry as depicted in Fig. 3(b). It was shown before[8,10] that the temperature effect of the wires is very localized since the temperature decreases to ambient values over a distance on the order of 100 μm. It is therefore presumable that this will only slightly influence the fluid flow around the heaters. As will be described in the next section, comparative simulations on the two infinite cylinders geometry show this assumption to be reasonable.

B. Simulation results

In the different simulations, the frequency f was varied between 0 and 10 kHz, and the magnitude u_0 was chosen as $3 \times 10^{-5} < u_0 < 1 \times 10^{-2}$ m/s (these values correspond in free space to sound levels of, respectively, 56 and 106 dB). Each simulation result provided the magnitude and phase of the particle velocity and the pressure at each point in space, such that the streamline pattern in the fluid could also be investigated. It was observed in the simulations that for the region of interest, $3 \times 10^{-5} < u_0 < 1 \times 10^{-2}$ m/s, all dynamics were linear in u_0, i.e., an increase of the amplitude of the imposed acoustic wave led to an equal increase of the velocities and pressures at all points in space.

For the two infinite cylinders geometry, we investigated first the influence of the presence of the small rectangular obstacle and of its temperature on the flow profile and on the magnitude of the particle velocity in between the cylinders.

The spacing between the cylinders was equal to 1.1 mm.
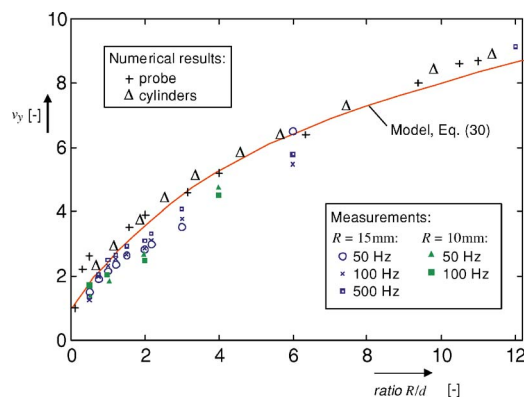
FIG. 9. (Color online) Particle velocity v_y at the point in the center between the two cylinders $[(x,y)=(0,0)]$, normalized by the imposed particle velocity at infinity, measured at different frequencies, with the sensor between two cylindrical objects. For comparison, the results of the computational analysis are shown, both on the commercial probe and on two infinite cylinders. For an air viscosity of $\nu = 1.5895 \times 10^{-5}$ m²/s, the theoretical curve following from Eq. (30) is also depicted, represented by the drawn line. The experimental results at 50 Hz were obtained by means of the shaker setup and therefore had to be scaled by a factor of 1.35 as described in Sec. IV; for 100 and 500 Hz the loudspeaker was used. Different cylinder radii (R =10 mm and R=15 mm) were investigated.

This spacing d corresponds to the effective distance between the cylinders of the commercial probe, which is the mutual cylinder spacing of 1.56 mm minus the thickness of the thin strip (a printed circuit board) on which the sensor is mounted (see Fig. 2).

When the temperature of the rectangular element was set to be T=700 K, a magnification of the particle velocity between the cylinders (with respect to the applied particle velocity u_0 of the incident sound wave) of 3.6±0.1 was found; when its temperature was the same as the ambient temperature, T=300 K, the amplification was 3.4±0.1. Apparently, the higher temperature results in only a slight increase of the particle velocity. The rectangular element itself had, taking into account the numerical error of the calculations, only a small influence on the magnitude of the particle velocity at the central point between the cylinders (3.5±0.1 and 3.4±0.1 for the situations with and without the element, respectively).

Next, the dimensions of the cylinders were varied. Keeping the mutual cylinder spacing d at a constant value of 1.1 mm, the cylinder radius R was varied in small steps from 1.1 to 12 mm. The same radii were also investigated at a mutual spacing of d=2.2 mm. The result of this, the magnification, i.e., the normalized particle velocity in between the cylinders, as a function of R/d is shown in Fig. 9. Also shown are the theoretical dependence on the ratio R/d, and measurement results described in Sec. IV.

The above-mentioned results show that both the effect of the sensor temperature and the effect of the presence of a small obstacle in the flow are relatively small. The following simulations, on the factual probe of specific interest, were therefore performed at constant temperature and without a representation of the sensor element, which reduced considerably the duration of the computations.

With respect to this specific geometry, we defined a number of points located in and around the probe to consider in particular. Point A is located in between the two small
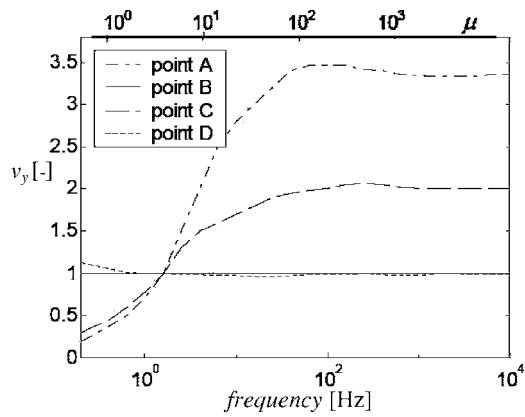
FIG. 10. Amplitude of the particle velocity at different points in and around the probe, normalized to the magnitude u_0 of the incoming wave. The upper horizontal axis shows the dimensionless frequency $\mu = 2\pi f l^2 / \nu$.

FIG. 11. Phase shift of the signals at points A, C, and D with respect to the phase at these points due to the imposed wave when there is no probe (i.e., the phase shift of A, C, and D due to the presence of the probe), as a function of frequency and μ. The lines are polynomial fits only to show the trends of the curves, they are not based on a model.

cylinders (where the sensor is placed in reality, but not in the simulation), point B is at a distance of 8 cm in front of the probe (on the outer boundary where the acoustic wave is imposed), C is located at 6.5 mm left from the center, and D was defined at 3 cm above A to investigate the phase behavior of the wave. See Fig. 3(b).

When the frequency rises, the particle velocity (normalized to the value u_0 of the imposed wave) increases at both points A and C, as shown in Fig. 10. One sees that, especially at A, a large amplification of the particle velocity is attained. This magnification approaches a value of about 3.3 at a frequency of 50 Hz. For frequencies below 1 Hz, however, the normalized particle velocity at A is smaller than 1, and decreases to 0.2 at 0.5 Hz. The same tendency is observed at C, a representative point for the region just next to the probe. The normalized signal increases from 0.3 at 0.5 Hz up to 1.7 at 10 Hz and remains almost constant above 100 Hz. With the perceived frequency behavior it seems natural to define a characteristic frequency to characterize the properties of this probe. Noting that at approximately $f = 1.5$ Hz both points A and C have a velocity amplification of 1, a reasonable choice seems to define this characteristic frequency of about $f_c = 1.5$ Hz, so that for frequencies below f_c, the (scaled) particle velocity amplitudes at A and C are lower than one, whereas for $f > f_c$, they increase.

The time dependence of the particle velocities at the points of most interest, A, C, and D, was analyzed in order to determine the local phase shift due to the presence of the probe (the difference between the phase of those points and the phase of those points if there were no probe at all, i.e., if the wave propagated in free space). Figure 11 shows this phase shift θ of A, C, and D, defined as $\theta_A = \phi_A - \phi_{A,\text{without any probe}}$. For high frequencies, θ_A and θ_C approach 0, while θ_D is, within the error margins, almost frequency independent and constant at -0.1 ± 0.1 rad. The curves in Fig. 11 are polynomial fits on the logarithmic frequency scale and are only depicted to show the trends of the curves.

In Sec. II it was mentioned the problem can be made nondimensional by scaling to a dimensionless frequency μ,

$$\mu = \frac{2\pi f l^2}{\nu}, \tag{41}$$

where l is a characteristic length of the geometry, for which we take the radius of the two small cylinders, and ν the kinematic viscosity of the fluid. In the original situation of the specific probe in air, we have $\nu_0 = 1.5895 \times 10^{-5}$ m^2/s and $l = R_0 = 2.47$ mm. Simulations were performed for different viscosities, varying in an extended range of $0 < \nu < 10\nu_0$, the case $\nu_0 = 0$ corresponding to a completely nonviscous gas. The probe itself was also scaled (with all dimensions scaling proportionally to l) in the range $R_0/2 < l < 10R_0$. It was found that, indeed, the relevant independent parameter for the problem is the scaled frequency μ. Further we verified that the simulated velocity values at all points in space were linear in u_0, the magnitude of the incoming sound wave, for values $u_0 < 10^{-2}$ m/s. The dependence of the particle velocity at the different points as a function of the dimensionless parameter μ is seen in Fig. 10. The velocity has been normalized by dividing through the velocity amplitude u_0 of the incoming wave at the boundary.

According to Fig. 10 it was noted that two different frequency regimes can be distinguished in which the flow profile is essentially different. For frequencies well below 1.5 Hz, the signal at A is strongly attenuated. Since the boundary layer thickness δ is in the order of $\delta \sim \sqrt{\nu/2\pi f}$, one finds that at 1 Hz, this layer is approximately equal to the mutual cylinder spacing. Therefore, point A is then situated in a large viscous boundary layer that extends over all space between the two cylinders. For frequencies much higher than 1.5 Hz, viscous effects become less and less important, the boundary layer thickness decreases, and the gas flows at A with increased particle velocity. Figure 12 is a visualization of this phenomenon that the particle velocity between the cylinders is attenuated for low frequencies and amplified for higher frequencies. For low frequencies ($f \sim 1$ Hz) the fluid tends to flow around the probe as a whole while the velocity in the boundary layer is low, as is seen in Fig. 12(a). This effect disappears for high frequencies.
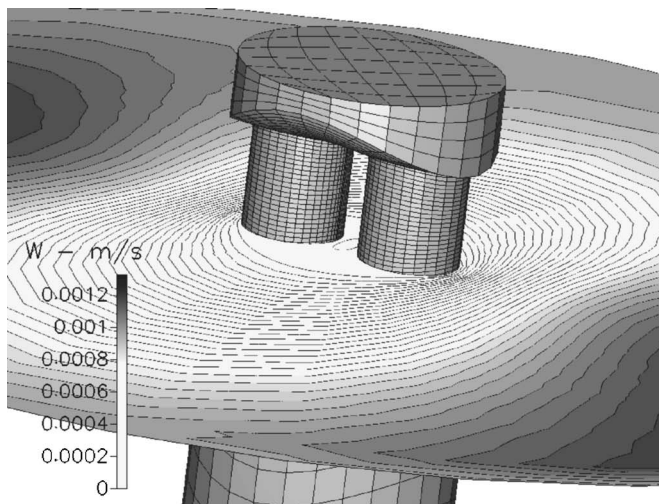
(a)
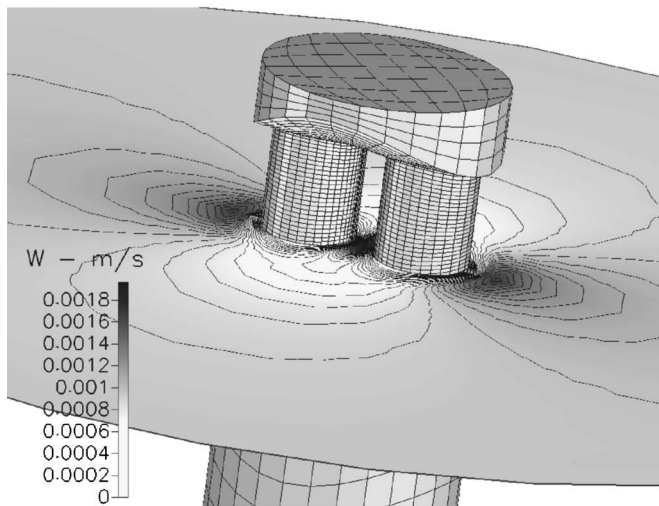
(b)

FIG. 12. (a) Contour plot of the particle velocity in a horizontal plane, for $f=0$ Hz. ($v_0=1$ mm/s). (b) As in (a) for $f=10$ Hz ($v_0=1$ mm/s).

FIG. 13. Normalized particle velocity v at point A, as a function of R/d, according to the computational analysis, with $f=10$ Hz and $\nu=1.5895 \times 10^{-5}$ m²/s.

The two different regimes can be distinguished also in terms of a characteristic value of the parameter μ, giving $\mu < \mu_c$ and $\mu > \mu_c$ with $\mu_c \approx 3.5$.

Next, we analyzed the dependence of the velocity amplification at A for high frequencies on the probe geometry. The dimensions of the two cylinders and their mutual distance were varied, with the mutual spacing between the cylinders d varying in the range $d_0/10 < d < 3d_0$ and the cylinder radius R changing from $R_0/2 < R < 10R_0$. The value R_0, the cylinder radius of the actual probe, was equal to 2.47 mm; for d_0 we took the effective gap distance between the cylinders, $d_0=1.1$ mm, corresponding to the mutual cylinder spacing minus the thickness of the strip the sensor is mounted on. Figure 13 shows this dependence of the normalized velocity at the central point between the cylinders, A, as a function of R/d, at a frequency of 10 Hz.

IV. EXPERIMENTAL WORK

To determine the effects of two cylindrical pillars on the output signal of the particle velocity sensor, a setup was re-

alized to compare two identical particle velocity sensors. One sensor had no adjacent cylinders and was used as a reference, while the other sensor was placed between two small cylindrical pillars of varying diameter. The setup consisted of two movable pillars of 5 cm height and diameters of 20 and 30 mm, allowing for a variation of the mutual pillar distance d. A second separate particle velocity sensor without adjacent cylinders was used as a reference sensor and placed at a relatively large distance, about 10 cm, from the pillars. A photograph of the measurement setup is shown in Fig. 14(a). The sound signal was generated by a loudspeaker acting as a point source that was placed at a distance of 0.5 m from the setup. The distance between the reference sensor and the cylinders pair should be as large as possible to avoid effects of this construction on the output signal like acoustic reflections. On the other hand, the sound field due to the sound source must be as similar as possible for both sensors. An acceptable compromise between these requirements was found at a distance between the sensors of 10 cm.

We performed a reference measurement without the two pillars mounted to determine the transfer function between the particle velocity sensor of investigation and the reference sensor. This transfer function had a value close to 10 in a frequency range of 100 Hz–1.5 kHz, indicating that in this frequency band no irregular effects occur. By varying the distance d between the pillars and the radius R of the pillars, the transfer function, i.e., the ratio of the output signal of the sensor and the reference sensor, was determined for a wide range of R/d values. The division by the reference measurement permitted one to eliminate any differences in the responses of the sensors so that merely the effects of the cylindrical obstructions were measured. Results of these measurements, in a frequency range of $100 < f < 1500$ Hz, are plotted in Fig. 14(b).

For the low-frequency response use was made of a "shaker" setup, providing a white noise excitation from 10 to 500 Hz. The movement of the sensor due to the shaker setup is assumed to yield a sensor output analogous to that due to a particle velocity of a sound wave. Similar to that described earlier, the cylinders pair with the central sensor, and a reference sensor, were mounted on the shaker and the transfer function was determined for varying R/d ratios. Not

van Honschoten et al.: Particle velocity sensor between cylindrical obstructions
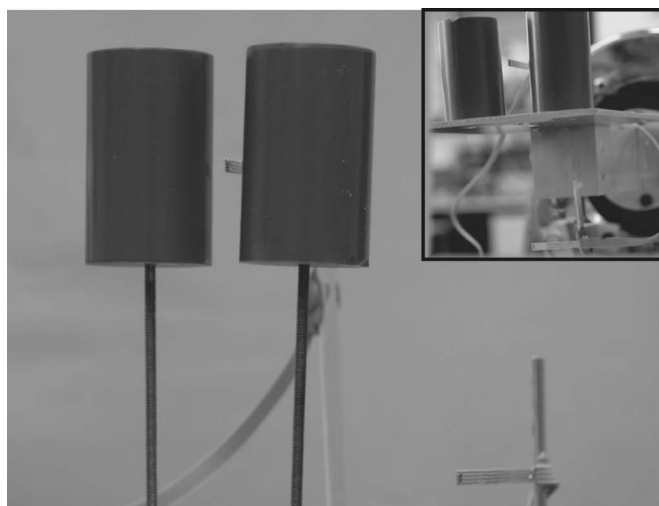
(a)
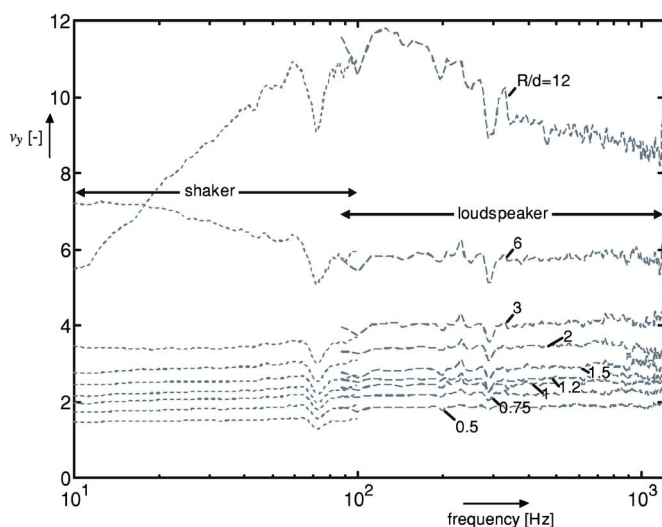
(b)

FIG. 14. (Color online) (a) Photograph of the experimental setup, showing the particle velocity sensor between the two small cylinders and the reference senor, mounted on a tripod. The inset shows the two cylinders mounted on the shaker platform. (b) Normalized velocity v_y in the center between the two cylinders [at $(x,y)=(0,0)$], as a function of frequency, with the ratio R/d as a parameter. For frequencies below 100 Hz, the measurements were performed by means of the "shaker" setup (with $R=15$ mm); for the higher frequencies a loudspeaker was used. Both sets of measurements are depicted by different symbols. The shaker measurements were all scaled by a factor of 1.35 to match optimally with the loudspeaker results for the higher frequencies.

only the mutual distance d was varied, also different cylinder dimensions ($R=10$ mm and $R=15$ mm) were investigated for this purpose. Although this experiment produced useful data for frequencies up to 500 Hz, the results for frequencies between 10 and 150 Hz were used.

The results for different frequencies were already shown in Fig. 9. As one can see in Fig. 9, for the frequencies $f=50$ Hz and $f=100$ Hz, the experimentally determined velocities lie somewhat below the simulation and theoretic results for ratios R/d around approximately 3. A possible explanation may be found in the fact that at relatively large cylinder spacings, the horizontal and vertical dimensions of the total object (two cylinders and their gap in between) are

on the same order of magnitude: The height h is not large anymore with respect to the other dimensions. This may lead to a flow over the structure as a whole, the gas tends to stream over what is perceived as a single obstacle. This may result in a decrease of the measured velocity at the sensor location.

For the current probe geometry of Fig. 2, i.e., the two sensor wires in between two cylinders of 4.94 mm diameter and an effective gap of 1.1 mm, yielding $R/d=2.2$, the amplification (this is the ratio of the output signal with / without probe package) was found to be 3.2 at 500 Hz. It must be noted that the precise value of this amplification is strongly dependent on the specific mounting of the particular probe under investigation.

Figure 14(b) illustrates the measurement results of the shaker setup, using pillars with a radius of $R=15$ mm, and the measurements by means of the loudspeaker together, as a function of frequency with R/d as a parameter. In the overlapping frequency range $85<f<100$ Hz the shaker and the loudspeaker results are plotted conjointly. The shaker measurements as they are depicted in Fig. 14(b) were scaled by a factor of 1.35 to correspond optimally to those of the loudspeaker. This division by 1.35 (the same factor for all shaker measurements) was required to match the two measurement series in the adjoining frequency ranges: ideally this scale factor is unity. Although the emergence of a difference between the shaker generated signal and that of the loudspeaker is not fully unexpected, the origin of this scale factor is not completely understood yet and is an important point of further investigation.

The dip at a frequency of about 70 Hz is attributed to a mechanical resonance of the shaker setup; it shows a frequency shift if the shaking platform is placed vertically instead of horizontally. For frequencies above 50 Hz the amplification is nearly independent of frequency for R/d ratios up to 12. At $R/d=12$ the low frequency "cut off" effect is manifest: A cut off frequency of approximately 50 Hz is observed. With $f=50$ Hz and $l\approx0.5$ mm, we find then for the dimensionless characteristic frequency, $\mu_c=2\pi fl^2/\nu\approx5$, which is close to the theoretical value of $\mu_c\approx4$.

V. CONCLUSIONS

We have analyzed the effects of two cylindrical objects adjacent to an acoustic sensor on the velocity profile around the sensor. This was done by a theoretical analysis on two parallel infinite cylinders, together with numerical simulations on the geometry. The theoretical two-dimensional analysis provided an analytical expression for the stream function around the cylinders. The supplemental computational simulations showed that both the presence of the small particle velocity sensor between the two cylinders and its high local temperature have only a small effect on the flow profile in and around the cylinders and on the particle velocity near the sensor. A large magnification of the particle velocity in between the cylinders for frequencies above 4 Hz was found in these simulations.

As a specific example of the two cylinder geometry, a commercially available probe was investigated experimen-

tally and by means of a computational analysis on this geometry. The magnification in this probe was found to increase up to approximately 3 in a frequency range from 50 up to 1000 Hz, being 3.2 at 500 Hz.

Although in these simulations the geometry differed from the ideal two-dimensional case of two cylinders, both the theory and the simulations on two cylinders are seen to give an adequate description for the fluid flow in the package of the probe. Therefore, with the two-dimensional model the so-called "package gain" at high frequencies can be described. For low frequencies, viscous effects dominate and signals are attenuated. This is important for acoustic measurement applications, in which low frequency signals should be suppressed and higher frequencies be amplified. In this respect one can define a characteristic value μ_c, of the dimensionless frequency $\mu = 2\pi f l^2 / \nu$, below which frequency acoustic signals are attenuated and above which they are amplified. Experimental results show an estimated value of $\mu_c \approx 5$, theory yields $\mu_c \approx 4$ where simulations give $\mu_c \approx 3.5$; a satisfactory agreement.

We found a relatively good correspondence among simulations, theory, and experiments. The two-dimensional analysis has provided a better understanding of the gas flow for low Reynolds numbers around parallel cylinders and since it is seen to be partly applicable for the specific probe too it opens the way, together with the numerical simulations, for further optimization of the probe geometry.

ACKNOWLEDGMENTS

[1] H. E. de Bree, P. J. Leussink, M. T. Korthorst, H. V. Jansen, T. S. J. Lammerink, and M. C. Elwenspoek, "The Microflown: A novel device measuring acoustical flows," Sens. Actuators, A **54**, 552–557 (1996).

[2] W. F. Druyvesteyn, H. E. de Bree, and M. Elwenspoek, *A New Acoustic Measurement Probe; The Microflown*, (IOA, London, 1999).

[3] H. Schurer, P. Annema, H. E. de Bree, C. H. Slump, and O. E. Herrmann, "Comparison of two methods for measurement of horn input impedance," Proceedings of the 100th AES convention, Copenhagen, 1996.

[4] F. J. M. van der Eerden, H. E. de Bree, and H. Tijdeman, "Experiments with a new acoustic particle velocity sensor in an impedance tube," Sens. Actuators, A **69**, 126–133 (1998).

[5] H. E. de Bree, T. Korthorst, P. J. Leussink, H. Jansen, and M. Elwenspoek, "A method to measure apparent acoustic pressure, flow gradient and acoustic intensity using two micromachined flow microphones," Eurosensors X, Leuven, 1996.

[6] H. E. de Bree, "Add-on Microflown for a high-end pressure gradient microphone," Proceedings of the 109th AES Convention, Los Angeles, 2000, No. 5181.

[7] H. E. de Bree, H. V. Jansen, T. S. J. Lammerink, G. J. M. Krijnen, and M. Elwenspoek, "Bi-directional fast flow sensor with a large dynamic range," J. Micromech. Microeng. **9**, 186–189 (1999).

[8] V. B. Svetovoy and I. A. Winter, "Model of the Microflown microphone," Sens. Actuators, A **86**, 171–181 (2000).

[9] J. W. van Honschoten, V. B. Svetovoy, G. J. M. Krijnen, and M. Elwenspoek, "Analytic model of a two-wire thermal sensor for flow and sound measurements," J. Micromech. Microeng. **14**, 1468–1477 (2004).

[10] J. W. van Honschoten, G. J. M. Krijnen, V. B. Svetovoy, H. E. de Bree, and M. Elwenspoek, "Optimisation of a two-wire thermal sensor for flow and sound measurements," Proc. MEMS 2001, Interlaken, Switzerland, pp. 523-526.

[11] Z. Carrière, "Double boundary layers in oscillatory viscous flow," J. Phys. Radium **10**, 673-677 (1929).

[12] E. N. Andrade, "On the circulations caused by vibrations of air in a tube," Proc. R. Soc. London, Ser. A **134**, 445-470 (1931).

[13] H. Schlichting, "Berechnung ebener periodischer Grenzschichtströmungen, (Calculation of periodical boundary layer flows)" Phys. Z. **23**, 327-335 (1932).

[14] N. Riley, "Oscillatory viscous flows," Mathematika **12**, 161-175 (1965).

[15] C. Y. Wang, "On high-frequency oscillatory flows," J. Fluid Mech. **32**, 55–68 (1968).

[16] Z. Zapryanov, Zh. Kozhoukharova, and A. Iordanova, "On the hydrodynamic interaction of two circular cylinders oscillating in a viscous fluid," J. Appl. Math. **39** 204-220 (1988).

[17] A. P. Zhuk, "A study of the interaction of an acoustic wave in a viscous liquid with two cylinders placed in parallel," Int. Appl. Mech. **27**, 321–327 (1991).

[18] R. Meneghini, F. Saltara, C. L. R. Siqueira, and J. A. Ferrari, "Numerical simulation of flow interference between two circular cylinders in tandem and side-by-side arrangements," J. Fluids Struct. **15**, 327–350 (2001).

[19] J. W. van Honschoten, D. R. Yntema, M. Dijkstra, V. B. Svetovoy, R. J. Wiegerink, and M. Elwenspoek, "Analysis of packaging effects on the performance of the microflown," Proceedings of the DTIP Conference of MEMS & MOEMS, Stresa, Italy, ISBN 2-916187-03-0, 2006.

[20] L. D. Landau and E. M. Lifschitz, *Fluid Mechanics*, Course of Theoretical Physics Vol. **6**, 2nd ed. (2003), Butterworth-Heinemann, Burlington, pp. 20-21.

[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, corrected and enlarged ed. (Academic, New York, 1980).

[22] R. W. Fox and A. T. McDonald, *Introduction to Fluid Mechanics*, 3rd ed. (John Wiley & Sons, New York, 1985).

[23] L. M. Milne-Thomson, *Theoretical Hydrodynamics* (Courier Dover, London, 1996).

[24] J. W. van Honschoten, W. F. Druyvesteyn, H. Kuipers, R. Raangs, and G. J. M. Krijnen, "Noise reduction in acoustic measurements with a particle velocity sensor by means of a cross-correlation technique," Acta. Acust. Acust. **90**, 394–355 (2004).

[25] CFD Research Corporation, "Technologies for engineering simulations," Huntsville, AL, www.cfdrc.com Accessed 4/9/2007.

[26] H. Versteeg and W. Malalasekra, *An Introduction to Computational Fluid Dynamics: The Finite Volume Method Approach* (Pearson Education Ltd., London, 1996).

Blind identification of aggregated microphones in time domain

Mitsuharu Matsumoto[a] and Shuji Hashimoto[b]

Department of Applied Physics, Waseda University, 55N-4F-10A, 3-4-1, Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

This paper introduces an algorithm for blind identification of aggregated microphones in time domain. The features of our approach are summarized as follows: (1) The proposed method treats the blind identification problem of anechoic mixtures in the time domain. (2) The proposed method can identify the gain of each microphone for the directions of sounds whose number is more than the number of the microphones. (3) The proposed method does not utilize the statistical independence of the sounds. The sounds may be not only statistically independent but may also be statistically dependent. (4) The proposed method treats the partially disjoint sounds in the time domain. The sounds may overlap in the frequency domain unlike the sparseness approach. (5) The proposed method does not need to estimate the intervals where sounds are disjoint. First, it is shown that the problem of blind identification and blind source separation can be described not as a convolutive model, but as an instantaneous model in the case of the anechoic mixing when aggregated microphones are assumed. The necessary conditions and the algorithm with experimental results are also described. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2714916]

I. INTRODUCTION

The microphone array plays an important role in sound signal processing.[1–5] It is applicable to hearing aids,[6] robot audition,[7–9] speech enhancement,[10] and so on. The conventional microphone array, known as *phased microphone array*, realizes the delay-sum type microphone array,[11] the adaptive microphone array,[12,13] and Direction of Arrival (DOA) estimation such as high resolution algorithms,[14–16] by utilizing the phase difference of each microphone. However, it is difficult to miniaturize the microphone array due to utilization of the phase difference of each microphone. Correspondingly, the authors proposed another type of microphone array that does not utilize the phase difference of each microphone but utilizes the directivities of microphones located at the same place, namely *aggregated microphones*.[17–19] In this method, all microphones are located in the same position, and directional microphones are arranged to differentiate the amplitude gains. Hence, the system can easily be miniaturized. This feature is useful when applied to small robots, conference systems, and so on.

The phased microphone array requires a precise estimation of the time difference from the sound to each microphone. In a similar way, the aggregated microphones method requires a precise estimation of the directivity of each microphone. The objective of our research is to simplify this process, that is, to identify a mixing parameter under the blind situation. Second-order blind identification (SOBI) is a blind identification algorithm that utilizes the second-order statistics of received sounds. SOBI is based on the unitary diagonalization of the whitened data covariance matrix.[20–22] There

also exist some blind identification algorithms that utilize higher order statistics.[23] Due to the calculation of higher order statistics, however, computational cost is large. On the other hand, in our method statistical information is not utilized to estimate the mixing matrix.

Some approaches utilize sparseness to solve the problem of blind identification.[24–26] Sparseness means that most of the frequency components of a signal are zero, so the sources rarely overlap in the frequency domain.[27,28] Under this assumption, it is possible to extract each signal using time-frequency binary masks. However, due to the binary masks, these approaches require that the frequency components of the mixed sounds hardly overlap at any time.

In this paper, we consider the problem of identifying the directivity of each microphone in the aggregated microphones by utilizing the nonstationary, partially disjoint signals, and propose a blind identification algorithm of aggregated microphones. Partially disjoint means the signals overlapping both in the time domain and the frequency domain while there exist some intervals where the sounds are disjoint. In our method, the mixing matrix is identified utilizing the intervals where the sounds are disjoint by utilizing the characteristics of the aggregated microphones.

We have already proposed a blind decomposition algorithm utilizing aggregated microphones.[29] However, it was only applicable to the case of two sounds and two microphones. This paper outlines an improved algorithm, that is, a generalized algorithm that is applicable to the case of multiple sounds and multiple microphones.

According to the paper surveyed by O'Grady *et al.*, signal mixtures of a microphone array are categorized as the instantaneous, anechoic, and echoic mixings.[30] In this paper, we discuss anechoic mixing. In Sec. II, we first explain the difference of a phased microphone array and the aggregated

---

[a] Electronic mail: matsu@shalab.phys.waseda.ac.jp
[b] Electronic mail: shuji@waseda.jp
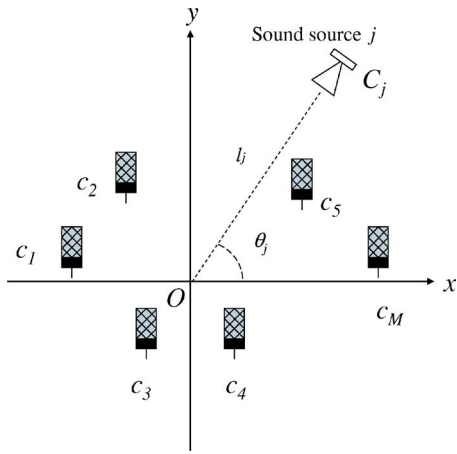
FIG. 1. (Color online) Basic concept of the phased microphone array.
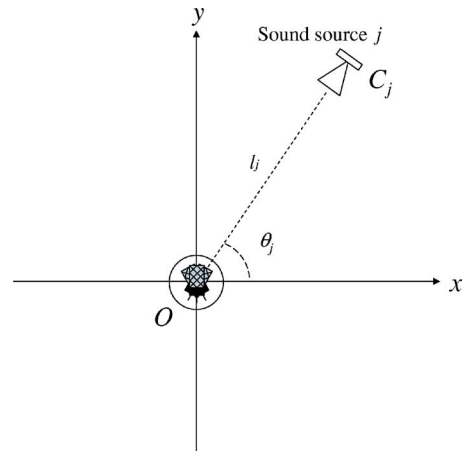
FIG. 2. (Color online) Basic concept of the aggregated microphones.

microphones and formulate the problem. We also show that the problem of blind identification and blind source separation can be described not as the convolutive model but as the instantaneous model in the case of anechoic mixing when aggregated microphones are assumed. In Sec. III, we explain the procedure of the proposed method, which can treat the problem not in the time-frequency domain but in the time domain due to the characteristics of the aggregated microphones. In other words, we can consider the problem not in complex space but in real space. The mixing matrix can be directly identified utilizing the observed signals without estimating the intervals where the signals are disjoint. In Sec. IV, we show the experimental results of blind identification and blind source separation. We show the proposed method can identify the directivity of each microphone for the directions of sounds whose number is not only equal to or less than the number of the microphones but also more than the number of microphones. We also show the proposed method can identify the mixing matrix of the aggregated microphones utilizing not only statistically independent signals but also statistically dependent signals with experimental results.

II. PROBLEM FORMULATION

Let us consider N sounds recorded by M microphones in an anechoic room. First, we will explain the difference between the phased microphone array and the aggregated microphones. Figures 1 and 2 illustrate the basic concept of the phased microphone array and the aggregated microphones, respectively. In Figs. 1 and 2, O represents the origin. The position of the ith microphone and the jth sound source are represented by c_i and C_j, respectively. The angle and the distance from the sound source j to the origin are represented by θ_j and l_j, respectively. In the phased microphone array, all the microphones are located at different positions to differentiate the phase. Omnidirectional microphones are usually utilized in order to make the amplitude gain identical. Hence, the received signal of the ith microphone $x_i(t)$ can be represented as follows:

$$x_i(t) = \sum_{j=1}^{N} s_j(t - \tau_{ij}) \quad (i = 1, 2, \ldots, M), \tag{1}$$

where $s_j(t)$ represents the sound j at the origin. The time delay from the origin to the ith microphone regarding the sound j is represented by τ_{ij}. In the phased microphone array, the delay τ_{ij} needs to be considered. Various applications such as the sound separation, DOA estimation, and so on, are realized utilizing the difference of the time delay τ_{ij}. On the other hand, in the aggregated microphones, directional microphones are utilized to differentiate the amplitude gains. All the microphones are located at the same positions to make the phase identical. Suppose all the microphones are located at the origin as shown in Fig. 2. The received signal $x_i(t)$ can be represented as follows:

$$x_i(t) = \sum_{j=1}^{N} d_{ij} s_j(t) \quad (i = 1, 2, \ldots, M), \tag{2}$$

where d_{ij} represents the amplitude gain of the ith microphone with regard to the sound j. In anechoic environments, signals at a sound source arrive at an observation point undistorted, but with an unknown time delay and attenuation in the case of phased microphone array. This time delay is a nonlinear parameter that can make finding solutions to blind identification and blind source separation (BSS) difficult. However, with the aggregated microphones, because the microphones are at the same location, time delays are no longer an issue. In other words, in the aggregated microphones, we can describe the blind identification problem of anechoic mixing as an instantaneous mixture unlike the phased microphone array. Various applications are realized utilizing the difference of the amplitude gain d_{ij} in the aggregated microphones. The phased microphone array requires information regarding the positions of all the microphones in advance because time delay is utilized when applying the phased microphone array to various applications. In a similar way, the aggregated microphones method requires information regarding the amplitude gain d_{ij} of all the microphones in advance. One of our research targets is to simplify this

M. Matsumoto and S. Hashimoto: Blind identification of aggregated microphones

process, that is, to identify d_{ij} for aggregated microphones.

To simplify the explanation, we rewrite $x_i(t)$ as follows:

$$x_i(t) = \sum_{j=1}^{N} D_{ij} s_j'(t) \quad (i = 1, 2, \ldots, M), \tag{3}$$

where

$$s_j'(t) = d_{1j} s_j(t). \tag{4}$$

D_{ij} represents the amplitude gain ratio of the ith microphone to the first microphone regarding the direction j as follows:

$$D_{ij} = \frac{d_{ij}}{d_{1j}}. \tag{5}$$

In this paper, we regard the blind identification problem as obtaining D_{ij} utilizing $x_i(t)$ except D_{1j} because $D_{1j}=1$.

III. BLIND IDENTIFICATION OF AGGREGATED MICROPHONES

A. Assumption regarding the sounds

Let us define $\Psi_{ii}(k,t)$ as the short-time autocorrelation function of the received sounds $x_i(t)$ at time t as follows:

$$\Psi_{ii}(k,t) = \frac{1}{L} \sum_{l=0}^{L-1} x_i(l+t) x_i(k+l+t), \tag{6}$$

where k and L represent the time deviation and the number of the data used for averages, respectively. We define the ratio of the interchannel power difference $\Delta A_i(t)$ as follows:

$$\Delta A_i(t) = \frac{\Psi_{ii}(0,t)}{\Psi_{11}(0,t)} \quad (i = 1, 2, \ldots, M). \tag{7}$$

We assume the following conditions regarding the sound sources.

Assumption 1: The signals are partially disjoint.

Let us define $\Phi_{jj}(k,t)$, the short-time autocorrelation function of $s_j(t)$ at time t as follows:

$$\Phi_{jj}(k,t) = \frac{1}{L} \sum_{l=0}^{L-1} s_j(l+t) s_j(k+l+t) \quad (j = 1, 2, \ldots, N). \tag{8}$$

Under these definitions, this condition can be described as follows: $\exists t_j (j = 1, 2, \ldots, N)$ such that

$$\Phi_{jj}(0, t_j) \neq 0,$$

$$\Phi_{mm}(0, t_j) = 0 \quad (m \neq j), \tag{9}$$

where $\Phi_{jj}(0,t)$ represents the short-time mean power of $s_j(t)$ at time t. Let us define the set T_j whose elements are t_j satisfying Eq. (9). Let us also define p_j as the number of the elements of T_j.

Assumption 2: Ununiformity of $\Delta A_i(t)$.

We assume that the multiple signals should vary when they occur simultaneously, so that the number of $\Delta A_i(t)$ values appearing the most frequently in the overlapping intervals is less than the number of the disjoint points. Overlapping intervals are the intervals where some or all of the signals overlap in the time domain. In general, $\Psi_{ii}(k,t)$ can be expressed as follows:

$$\Psi_{ii}(k,t) = \sum_{j=1}^{N} D_{ij}^2 \Phi_{jj}(k,t) + \sum_{m=1}^{N} \sum_{n=1, n \neq m}^{N} D_{im} D_{in} \Phi_{mn}(k,t), \tag{10}$$

where

$$\Phi_{mn}(k,t) = \frac{1}{L} \sum_{l=0}^{L-1} s_m(l+t) s_n(k+l+t). \tag{11}$$

The first summation of Eq. (10) represents the summation of the autocorrelation of N sounds. The second summation of Eq. (10) represents the summation of the cross correlation of N sounds. $\Delta A_i(t)$ can be expressed as follows:

$$\Delta A_i(t) = \frac{\displaystyle\sum_{j=1}^{N} D_{ij}^2 \Phi_{jj}(0,t) + \sum_{m=1}^{N} \sum_{n=1, n \neq m}^{N} D_{im} D_{in} \Phi_{mn}(0,t)}{\displaystyle\sum_{j=1}^{N} \Phi_{jj}(0,t) + \sum_{m=1}^{N} \sum_{n=1, n \neq m}^{N} \Phi_{mn}(0,t)}. \tag{12}$$

Let us define q_i as the maximal number such that $\Delta A_i(t)$ has the same value in spite of the mixtures of multiple sounds. Assumption 2 can be expressed as follows:

$$p_j > q_i. \tag{13}$$

We will verify that speech signals satisfy this assumption enough to allow for blind identification of the aggregated microphones later. In a binary mask, it should be assumed that the frequency components of the mixed sounds hardly overlap in any time. On the other hand, in our method, the frequency components of mixed sounds may overlap in most of the frequency domain. The mixed sounds may also overlap in the time domain. Moreover, the sounds may not only be statistically independent but may also be statistically dependent. It should be noted that we do not need to know the intervals where the sounds are disjoint in the time domain through all the procedure.

B. Blind identification algorithm

Under Assumptions 1 and 2, D_{ij} can be identified as follows:

Procedure 1: Check $\Delta A_i(t)$ for each i.

Figure 3 depicts the basic concept of the time fluctuation regarding $\Delta A_i(t)$. As shown in Fig. 3, $\Delta A_i(t)$ is constant when only one sound is present, while $\Delta A_i(t)$ varies when multiple sounds are present under Assumptions 1 and 2. According to Assumption 1, $\Delta A_i(t_j)$ can be expressed as follows:

$$\Delta A_i(t_j) = \frac{D_{ij}^2 \Phi_{jj}(0,t_j)}{D_{1j}^2 \Phi_{jj}(0,t_j)} = \frac{D_{ij}^2 \Phi_{jj}(0,t_j)}{\Phi_{jj}(0,T_j)} = c_{ij}, \tag{14}$$

where c_{ij} is a constant value. Equation (9) is satisfied by t_j. If only one source $s_j(t)$ occurs in a short time L as shown in Eq. (14), $\Delta A_i(t_j)$ is constant and equal to c_{ij}, while $\Delta A_i(t)$ usually takes different values if multiple sounds are present and if
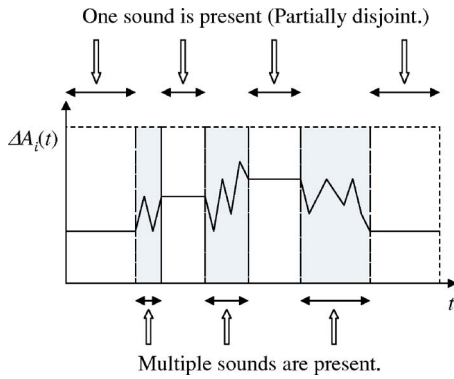
FIG. 3. (Color online) Basic concept of $\Delta A_1(i)$.

FIG. 4. The histogram of $\Delta A_2(t)$.

Assumption 2 is satisfied as shown in Fig. 3. When we rank $\Delta A_i(t)$ in descending order under Assumptions 1 and 2, $c_{ij}(j=1,2,\ldots,N)$ is ranked in the top N. Hence, we can obtain the estimated D_{ij} by utilizing the value of $\Delta A_i(t)$ ranked in the top N as follows:

$$\overline{D_{ij}} = \sqrt{\Delta A_i(t)}, \tag{15}$$

where $\overline{D_{ij}}$ represents either of the estimated values of D_{ij} except the order of D_{ij}.

Procedure 2: Sort $\overline{D_{ij}}$ to correspond to D_{ij}.

The order of $\overline{D_{ij}}$ depends on the characteristics of the sounds, that is, it does not correspond to the order of D_{ij}. We need to sort the columns of $\overline{D_{ij}}$ as the order of $\overline{D_{ij}}$ is the same as the order of D_{ij} in order to obtain the directivity for each direction. To solve this problem, we should check the number of t_k such as

$$\Delta A_i(t_k) = c_{ik}, \tag{16}$$

$$\Delta A_j(t_k) = c_{jk}. \tag{17}$$

We can regard $\overline{D_{ik}}$ and $\overline{D_{jk}}$ as the elements of the same column where the number of t_k is maximal. This process is similar to permutation in ICA. Note that Procedure 1 is executed independently for each i, that is, D_{ij} can be identified even if $M < N$. In this procedure, we do not need to estimate even the intervals where the sounds are disjoint in order to identify D_{ij}. We only have to check the values of $\Delta A_i(t)$ for a certain time, which include the intervals where the signal is disjoint. This method works out due to the characteristics of the aggregated microphones. This is because autocorrelation function only has complete information about the spectral amplitude and not the phase. Note that we do not execute this operation when $\Psi_{ii}(0,t)$ are 0.

IV. EXPERIMENT

A. Blind identification in the case that $M < N$

To evaluate the proposed method, we first conduct the simulation experiments regarding blind identification with $M=2$, $N=3$, that is, in the case that $M<N$. In our formulation, as we consider sounds that are partially disjoint, the BSS problem is the ill-posed problem in the case that $M<N$. However, in the proposed method, it is possible to identify D_{ij} even under such conditions. To obtain the gains of the microphones, we measured the gains experimentally. We utilized two directional capacitor microphones (ECM-TS125, Sony). The sounds are generated from 0°, 90°, and 180°. The microphones are set apart from 0.5 m. Based on the gains obtained in the experiment, we set the mixing matrix $\mathbf{D}=[D_{ij}]$ as follows:
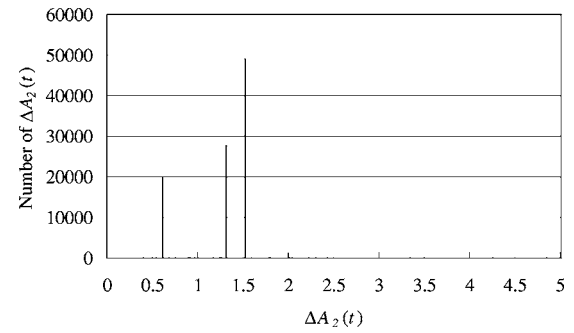
$$\mathbf{D} = \begin{bmatrix} 1 & 1 & 1 \\ 1.31 & 0.61 & 1.52 \end{bmatrix}. \tag{18}$$

We conducted eight experiments utilizing the mixed sounds. As the sound sources, we utilized "Japanese Newspaper Article Sentences" edited by the Acoustical Society of Japan. To clarify the difference between ICA and the proposed methods, we conducted the experiments utilizing not only the different signals but also the same signals with a time delay, that is, signals that are strongly dependent. In this experiment, we set L to 1 and the resolution of D_{ij} to 0.01 intervals. Figure 4 depicts one of the histograms of $\Delta A_2(t)$. The values of $\Delta A_2(t)$ ranked in the top 3 are 1.52, 1.31, and 0.61, that is, the D_{23}, D_{21}, and D_{22} values as shown in Fig. 4. We could also obtain similar histograms utilizing other mixed signals. The reliability of the estimated values depends on the number of the disjoint intervals in the time domain. If there is a sufficient number of disjoint intervals, the estimated values are relatively reliable. Table I shows the dominant three $\Delta A_2(t)$ values that appear frequently. The results of Experiments 1–4 in Table I show the results when different sounds are utilized. The results of Experiments 5–8 in Table I show the results when the same sounds are utilized, that is, the sounds are strongly dependent. In both cases, we can obtain the correct D_{21}, D_{22}, and D_{23} values as shown in Table I, while the order of D_{21}, D_{22}, and D_{23} depends on the characteristics of the sounds.

TABLE I. Dominant three $\Delta A_2(t)$.

Order	1	2	3
Experiment 1	1.31	0.61	1.52
Experiment 2	1.52	1.31	0.61
Experiment 3	0.61	1.52	1.31
Experiment 4	0.61	1.52	1.31
Experiment 5	1.52	1.31	0.61
Experiment 6	1.52	1.31	0.61
Experiment 7	1.52	1.31	0.61
Experiment 8	1.51	0.61	1.31

TABLE II. Dominant three $\Delta A_i(t)$ before sorting.

Order	1	2	3
Experiment 1($\overline{D_{2j}}$)	0.43	0.85	2
Experiment 1($\overline{D_{3j}}$)	1.31	0.61	1.52
Experiment 2($\overline{D_{2j}}$)	2	0.43	0.85
Experiment 2($\overline{D_{3j}}$)	1.52	1.31	0.61
Experiment 3($\overline{D_{2j}}$)	0.85	2	0.43
Experiment 3($\overline{D_{3j}}$)	0.61	1.31	1.52
Experiment 4($\overline{D_{2j}}$)	0.85	2	0.43
Experiment 4($\overline{D_{3j}}$)	0.61	1.31	1.52
Experiment 5($\overline{D_{2j}}$)	2	0.43	0.85
Experiment 5($\overline{D_{3j}}$)	1.52	1.31	0.61
Experiment 6($\overline{D_{2j}}$)	2	0.43	0.85
Experiment 6($\overline{D_{3j}}$)	1.52	1.31	0.61
Experiment 7($\overline{D_{2j}}$)	2	0.43	0.85
Experiment 7($\overline{D_{3j}}$)	1.52	1.31	0.61
Experiment 8($\overline{D_{2j}}$)	2	0.85	0.43
Experiment 8($\overline{D_{3j}}$)	1.51	0.61	1.31

TABLE III. Dominant three $\Delta A_i(t)$ after sorting.

Order	1	2	3
Experiment 1($\overline{D_{2j}}$)	0.43	0.85	2
Experiment 1($\overline{D_{3j}}$)	1.31	0.61	1.52
Experiment 2($\overline{D_{2j}}$)	2	0.43	0.85
Experiment 2($\overline{D_{3j}}$)	1.52	1.31	0.61
Experiment 3($\overline{D_{2j}}$)	0.85	2	0.43
Experiment 3($\overline{D_{3j}}$)	0.61	1.52	1.31
Experiment 4($\overline{D_{2j}}$)	0.85	2	0.43
Experiment 4($\overline{D_{3j}}$)	0.61	1.51	1.31
Experiment 5($\overline{D_{2j}}$)	2	0.43	0.85
Experiment 5($\overline{D_{3j}}$)	1.52	1.31	0.61
Experiment 6($\overline{D_{2j}}$)	2	0.43	0.85
Experiment 6($\overline{D_{3j}}$)	1.52	1.31	0.61
Experiment 7($\overline{D_{2j}}$)	2	0.43	0.85
Experiment 7($\overline{D_{3j}}$)	1.52	1.31	0.61
Experiment 8($\overline{D_{2j}}$)	2	0.85	0.43
Experiment 8($\overline{D_{3j}}$)	1.51	0.61	1.31

B. Blind separation in the case that $M=N$

We also conducted BSS simulation experiments. As is well known, we can separate the sounds in the case that $M \geq N$ if D_{ij} can be estimated. To evaluate the proposed method, we also conducted experiments regarding blind source separation when $M=3$, $N=3$, that is, in the case that $M=N$. To obtain the gains of the microphones, we also measured the gains experimentally. We utilized three directional capacitor microphones (ECM-TS125, Sony). The sounds are generated from 0°, 90°, and 180°. The microphones are set apart from 0.5 m. Based on the gains obtained, we set the mixing matrix $\mathbf{D}=[D_{ij}]$ as follows:

$$\mathbf{D} = \begin{bmatrix} 1 & 1 & \\ 1.31 & 0.61 & 1.52 \\ 0.43 & 0.85 & 2 \end{bmatrix}. \tag{19}$$

We conducted eight experiments utilizing mixed sounds. As sound sources, we utilized "Japanese Newspaper Article Sentences" edited by the Acoustical Society of Japan. To clarify the difference between ICA and the proposed methods, we conducted the experiments utilizing not only the different signals but also the same signals with a time delay, that is, signals that are strongly dependent. Table II shows the dominant three $\Delta A_2(t)$ values before sorting $\overline{D_{ij}}$. The rows named $\overline{D_{2j}}$ and $\overline{D_{3j}}$ in Table II represent the values of $\overline{D_{2j}}$ and $\overline{D_{3j}}$ of each experiment, respectively. The results of Experiments 1–4 in Table II show the results when different sounds are utilized. The results of Experiments 5–8 in Table II show the results when the same sounds are utilized. In both cases, we can obtain the correct $\overline{D_{ij}}$ as shown in Table II, while the order of $\overline{D_{ij}}$ depends on the characteristics of the sounds. In separating the sounds, we need to estimate not only $\overline{D_{ij}}$ but also the order of $\overline{D_{ij}}$. Table III shows the dominant three $\Delta A_2(t)$ values after sorting $\overline{D_{ij}}$ utilizing Procedure 2. As well as the results in Table II, the results of Experiments 1–4 in Table III show the results in the case of utilizing the different sounds. The results of Experiments 5–8 in Table III show the results in the case of utilizing the same sounds. We can obtain the correct order of the columns of $\overline{D_{ij}}$ as shown in Table III. If the time delay is ignorable, the separation process can also be executed in the time domain. In the simulation experiment, we could obtain the complete waveform of the signals as separated signals.

C. Blind separation utilizing data recorded in the real world

To prove the effectiveness of the proposed method in a real environment, we also conducted the experiments, utilizing data recorded in the real world in the case of $M=2$ and $N=2$. In the experiments, we set the microphones as shown in Fig. 5. We utilized two directional capacitor microphones (ECM-TS125, Sony). The frontal direction of the one directional microphone is set as 90°. The frontal direction of the other directional microphone is set as 210°. Sounds are generated from 0° and 90°. The process was performed for sounds sampled with a sample frequency of 11050 Hz and 16 bit resolution. The microphones are set apart from 0.5 m. According to the experiments, the correct of $\mathbf{D}=[D_{ij}]$ is obtained as follows:
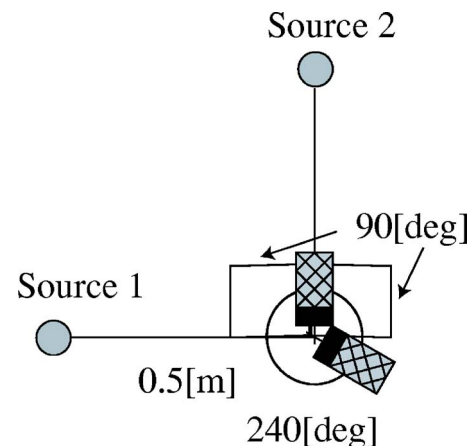
FIG. 5. (Color online) Arrangement for the experiments under the real world.

TABLE IV. Dominant two $\Delta A_2(t)$ utilizing data recorded in the real environment.

Order	1	2
Experiment 1	1.3	0.6
Experiment 2	0.6	1.3
Experiment 3	0.6	1.3
Experiment 4	0.6	1.3
Experiment 5	0.6	1.3
Experiment 6	0.6	1.3
Experiment 7	0.6	1.3
Experiment 8	0.6	1.3

FIG. 6. NRR corresponding to the value of D_{21}.

$$\mathbf{D} = \begin{bmatrix} 1 & 1 \\ 1.31 & 0.61 \end{bmatrix}. \tag{20}$$

Through these experiments in the real world, we discovered the following problems.

(1) When the estimated values include some error that is more than the size of quantization, similar values are ranked in the top N as different values. This problem is considered to be resoluble by enlarging the size of quantization.

(2) There was some phase difference in the real world case even when all the microphones were located at the same place, due to the microphones' phase responses and the size of the microphones. This problem is considered to be resoluble by enlarging the size of L.

To solve these problems, we set L to 256 and resolution of D_{ij} to 0.1 intervals. Table IV shows the dominant two $\Delta A_2(t)$ values. To evaluate the effectiveness of the proposed method for the dependent signals, we conducted the experiments utilizing not only the different sounds, but also the same sounds with time delay. The results of Experiments 1–4 show the results when different sounds are utilized. The results of Experiments 5–8 show the results in the case of utilizing the same sounds, that is, the sounds are strongly dependent. There will be some phase difference in the real world case even if all the microphones are located at the same place due to the microphones phase responses and the size of the microphones. In spite of the phase difference, however, we could estimate the gain of the aggregated microphones when the resolution of D_{ij} is 0.1 interval. The proposed method basically aims not at separating the sounds but at estimating the gains of the microphones for the individual sounds. However, as a result, when $M \geq N$ we can also separate the sounds. To evaluate the performance of the separation, we also calculate Noise Reduction Ratio (NRR). In this experiment, as we have two sounds, we regarded one sound as signal and the other sound as noise. Let $s(t)$ and $n(t)$ be the signal and the noise, respectively. We calculate Signal to Noise Ratio (SNR) as follows:
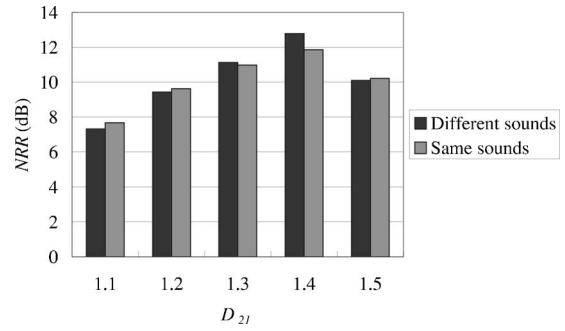
$$\text{SNR} = 10 \log_{10} \frac{\sum_{t=0}^{T} s^2(t)}{\sum_{t=0}^{T} n^2(t)}, \tag{21}$$

where T represents the time length. NRR is defined as follows:

$$\text{NRR} = \text{SNR}_{\text{after}} - \text{SNR}_{\text{before}}, \tag{22}$$

where $\text{SNR}_{\text{after}}$ is the SNR after signal separation. $\text{SNR}_{\text{before}}$ is that before separation. To confirm the robustness of the estimation error of D_{ij}, we calculated NRR values corresponding to the values of D_{ij}. Due to a tiny phase difference among sounds recorded by the microphones in the real world, the separation results in the time-frequency domain were much better than those in the time domain. Hence, we show the results in the case of executing the separation process in the time-frequency domain. To separate the sounds in the time-frequency domain, we first transform $x_i(t)$ into frequency components $X_i(\omega, t)$ by the short time discrete Fourier transformation (DFT). The frequency is represented by ω. The power of the separated signal vector $\mathbf{S}'(\omega, t) = [S_j'(\omega, t)]$ can be obtained as follows:

$$|\mathbf{S}'(\omega, t)| = \bar{\mathbf{D}}^{-1}|\mathbf{X}(\omega, t)|, \tag{23}$$

where $\mathbf{X}(\omega, t) = [X_i(\omega, t)]$ represents the received vector in the time-frequency domain. $X_j(\omega, t)$ represents the frequency components of $x_i(t)$. $\bar{\mathbf{D}} = [\overline{D_{ij}}]$ represents the estimated mixing matrix. The separated signal $s_j'(t)$ can be obtained by utilizing $S_j'(\omega, t)$. The phase of $X_1(\omega, t)$ was added as well as the spectral subtraction when the frequency components $S_j'(\omega, t)$ are transformed into the time components by inverse short time DFT. The black bar and gray bar in Fig. 6 show the average NRR corresponding to the value of D_{21} when we utilized four pairs of different sounds and the same sounds, respectively. In this case, we regarded $s_2(t)$ as the signal, while we regarded $s_1(t)$ as the noise. The black bar and gray bar in Fig. 7 show the average NRR corresponding to the value of D_{22} when we utilized four pairs of different sounds and the same sounds. In this case, we regarded $s_1(t)$ as the signal, while we regarded $s_2(t)$ as the noise. As shown in Figs. 6 and 7, NRR is improved even when D_{21} and D_{22} have some estimation error. In other
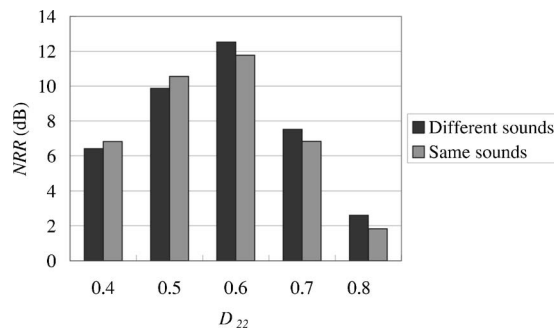
FIG. 7. NRR corresponding to the value of D_{22}.

words, the performance of the separation is relatively robust for the estimation error of the D_{ij} when the values of the amplitude gain are sufficiently different. However, it is considered that the error is relatively larger when the amplitude gains are relatively similar as well as that the results of the phased microphone array get worse when the time delay of the microphones is similar.

V. DISCUSSION AND CONCLUSION

In this paper, we proposed a method to identify the mixing matrix of aggregated microphones utilizing the sounds that are partially disjoint based on the gain difference among multiple microphones located in the same place. The proposed method can identify the mixing matrix of the aggregated microphones not only in the case that the number of the sounds is equal to or less than the number of the microphones but also in the case that the number of the sounds is more than the number of the microphones. The sounds may overlap in most of the time domain and the frequency domain unlike in the sparseness approach. The sounds may also be not only statistically independent but also dependent. We need to know neither the directivity of each microphone nor the intervals where the sounds are disjoint in advance. We have shown experimental results under simulated conditions and in real environments. As a result, we could obtain estimated values not only in the case of utilizing the different sounds, but also in the case of utilizing the same sounds. The proposed method, however, sometimes could not estimate the values of gains, for some of the following reasons:

(1) When the estimated values include errors that are more than the size of quantization, the values are estimated as the different values. This problem is considered to be resoluble by enlarging the size of quantization to some extent.
(2) When the microphones are not located sufficiently close to each other, the estimated errors tend to increase. This problem is considered to be resoluble by enlarging the size of L.
(3) When the length of the disjoint time of one sound is much larger than the disjoint time of the other sound, the estimated values are estimated as different values. This problem is considered to be resoluble by observing the sounds for sufficient time to some extent.

In future work, we would like to extend the proposed method for the case of the echoic environment. We are also considering to combine this method and the conventional aggregated microphones method, which utilizes the knowledge of the directivity of each microphone.

ACKNOWLEDGMENTS

[1]K. Sasaki and K. Hirata, "3D-localization of a stationary random acoustic source in near-field by using 3 point-detectors," Trans. CSICE **34**, 1329–1337 (1998).
[2]Y. Yamasaki and T. Itow, "Measurement of spatial information in sound fields by the closely located four point microphone method," J. Acoust. Soc. Jpn. **10–2**, 101–110 (1990).
[3]J. Huang, N. Ohnishi, and N. Sugie, "Spatial localization of sound sources: Azimuth and elevation estimation," Proceedings of IEEE Instrumentation and Measurement Technology Conference, 1998, pp. 330–333.
[4]H. F. Silverman, "An algorithm for determining talker location using a linear microphone array and optical hyperbolic fit," Proceedings of the Speech and Natural Language Workshop, 1990, pp. 151–156.
[5]S. Azuma, S. Uchikoshi, and K. Kido, "Studies on the spatial distribution of sensitivity in arc-arrayed microphone system," J. Acoust. Soc. Jpn. **40–10**, 677–683 (1985).
[6]A. Wang, K. Yao, R. E. Hudson, D. Korompis, and F. Lorenzelli, "Microphone array for hearing aid and speech enhancement applications," International Conference on Application-Specific Systems, Architectures, and Processors, 1996, pp. 231–239.
[7]T. Nishiura, M. Nakamura, A. Lee, H. Saruwatari, and K. Shikano, "Talker tracking display on autonomous mobile robot with a moving microphone array," Proceedings of the International Conference on Auditory Display, 2002, pp. 244–247.
[8]J.-M. Valin, J. Rouat, and F. Michaud, "Enhanced robot audition based on microphone array source separation with post-filter," Proceedings of the IEEE/RSJ International Conference on Robots and Intelligent Systems, 2004, pp. 2123–2128.
[9]J.-M. Valin, F. Michaud, B. Hadjou, and J. Rouat, "Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach," IEEE International Conference on Robotics and Automation, 2004, pp. 1033–1038.
[10]H. Luts, J. Maj, W. Soede, and J. Wouters, "Better speech perception in noise with an assistive multi-microphone array for hearing aids," Ear Hear. **25**, 411–420 (2004).
[11]K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura, and J. Kojima, "A microphone array system for speech recognition," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997, pp. 215–218.
[12]Y. Kaneda and J. Ohga, "Adaptive microphone array system for noise reduction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**, 1391–1400 (1986).
[13]K. Takao, M. Fujita, and T. Nishi, "An adaptive antenna array under directional constraint," IEEE Trans. Antennas Propag. **24**, 662–669 (1976).
[14]J. Capon, "High resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 2408–2418 (1969).
[15]S. S. Reddi, "Multiple source location: A digital approach," IEEE Trans. Aerosp. Electron. Syst. **AES-15**, 95–105 (1979).
[16]R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag. **AP-34**, 276–280 (1986).

[17] M. Matsumoto and S. Hashimoto, "Multiple signal classification by aggregated microphones," IEICE Trans. Fundamentals **E88-A**, 1701–1707 (2005).

[18] M. Matsumoto and S. Hashimoto, "Minimum variance method by aggregated micro-phones," Proceedings of the International Symposium on Signal Processing and Its Applications 2005, pp. 879–882.

[19] M. Matsumoto and S. Hashimoto, "A miniaturized adaptive microphone array under directional constraint utilizing aggregated microphones," J. Acoust. Soc. Am. **119**, 352–359 (2006).

[20] A. Belouchrani, K. Abed Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," IEEE Trans. Signal Process. **45**, 434–444 (1997).

[21] D. Nuzillard and J. M. Nuzillard, "Second order blind source separation on the Fourier space of data," Signal Process. **83**, 627–631 (2003).

[22] K. I. Diamantaras, A. P. Petropulu, and B. Chen, "Blind two-input-two-output FIR channel identification based on frequency domain second-order statistics," IEEE Trans. Signal Process. **48**, 534–542 (2000).

[23] B. Chen and A. P. Petropulu, "Frequency domain blind MIMO system identification based on second- and higher-order statistics," IEEE Trans. Signal Process. **49**, 1677–1688 (2001).

[24] T. Ihara, M. Handa, T. Nagai, and A. Kurematsu, "Multi-channel speech separation and localization by frequency assignment," IEICE Trans. Fundamentals **J86-A**, 998–1009 (2003).

[25] S. Rickard and O. Yilmaz, "On the approximate w-disjoint orthogonality of speech," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2002, pp. 529–532.

[26] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," IEEE Trans. Signal Process. **52**, 1830–1847 (2004).

[27] M. Aoki, M. Okamoto, S. Aoki, H. Matsui, T. Sakurai, and Y. Kaneda, "Sound source segregation based on estimating incident angle of each frequency component of input sources acquired by multiple microphones," Acoust. Sci. & Tech. **22**, 149–157 (2001).

[28] M. Aoki, Y. Yamaguchi, K. Furuya, and A. Kataoka, "Modified SAFIA: Separation of target signal close to the microphones and noise signal far from the microphones," IEICE Trans. Fundamentals **E88-A**, 468–479 (2005).

[29] M. Matsumoto and S. Hashimoto, "Blind source separation of anechoic mixtures in time domain utilizing aggregated microphones," European Signal Processing Conference, 2006 (CD-Rom).

[30] P. D. O'Grady, B. A. Pearlmutter, and S. T. Rickard, "Survey of sparse and non-sparse methods in source separation," Int. J. Imaging Syst. Technol. **15**, 18–33 (2005).

Multiaspect target detection via the infinite hidden Markov model

Kai Ni, Yuting Qi, and Lawrence Carin[a)]

Electrical and Computer Engineering Department, Duke University, Durham, North Carolina 27708-0291

A new multiaspect target detection method is presented based on the infinite hidden Markov model (iHMM). The scattering of waves from a target is modeled as an iHMM with the number of underlying states treated as infinite, from which a full posterior distribution on the number of states associated with the targets is inferred and the target-dependent states are learned collectively. A set of Dirichlet processes (DPs) are used to define the rows of the HMM transition matrix and these DPs are linked and shared via a hierarchical Dirichlet process. Learning and inference for the iHMM are based on a Gibbs sampler. The basic framework is applied to a detailed analysis of measured acoustic scattering data. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2714912]

I. INTRODUCTION

In many acoustic sensing scenarios the target is observed from multiple target-sensor orientations (or aspects), and the underlying acoustic scattered wave forms are highly aspect-dependent. It is often difficult to achieve reliable identification based on a single target view, because the wave forms emitted from two different targets may be similar at certain angles and easily confused. This motivates using a sequence of aspect-dependent looks at a single target, with this *sequential* information offering the potential to substantially improve identification performance.

The data observed when sensing a target from T target-sensor orientations are represented as a sequence of T scattered wave forms. Each of the wave forms is typically represented in terms of a set of features, and in the work presented here physics-based matching-pursuits feature extraction[1–3] is performed. Hidden Markov models (HMMs) have been successfully applied to model such sequential data for target detection and classification problems.[1,2,4] Each HMM "state" represents a set of generally contiguous target-sensor orientations over which the signal statistics are relatively invariant. The T feature vectors characteristic of the sequential data implicitly sample a sequence of T target states. The probability of transitioning from one state to another on consecutive measurements is modeled as a Markov process. Furthermore, because the target is usually concealed or distant, the target-sensor orientation is unknown and the actual sampled state sequence is hidden, motivating an HMM.

In the context of target detection and classification using HMMs, a key issue is to develop a methodology for defining an appropriate set of states. In some problems the state decomposition may be inferred from the underlying physics. For example, acoustic scattering from an elastic target is typically categorized by diffractive,[5–7] resonant,[8] and chirp-like (dispersive) scattering;[9] hence, ideally an HMM state may be defined by a set of angles for which a particular class of physics dominates. However, in many problems it is not possible to easily find a relationship between the physics and the underlying state. In previous work the state decomposition has been performed in an *ad hoc* manner,[1,2,4] requiring trial and error to manually select the model structure, (e.g., number of states). In the work reported here we investigate the idea of an *infinite* hidden Markov model (iHMM), which by construction has an infinite number of hidden states. We use a set of Dirichlet processes (DPs)[10] to build the iHMM. The Dirichlet process is often used to define mixture models with an infinite number of components,[11,12] but the simple DP form is insufficient for an iHMM; in an iHMM each row of the transition matrix is modeled as a DP and multiple rows are linked because they share the same set of hidden states. Therefore we use a hierarchical Dirichlet process (HDP) to constitute the iHMM. Proposed by Teh *et al.*,[13,14] the HDP is a nonparametric Bayesian model for sharing clusters among related groups. However, the HDP assumes a fixed partition of groups of data, while the group partition is random in an iHMM. In addition, the iHMM must utilize the assumed underlying temporal information of the sequential data, which is not considered in the original HDP.

Another limitation of many target detection and classification algorithms is that they usually require substantial training data, assumed to be similar to the data on which the algorithm is tested. Unfortunately, in many sensing applications one often has limited training data. In acoustic target detection problems, for instance, one may have multiple sets of limited scattering data, with each data set collected from different but related targets. Rather than building models for each target individually, as adopted in Refs. 1, 2, and 15, it is desirable to appropriately share the information among these related targets, thus offering the potential to improve overall identification performance. The iHMM effectively solves this problem using the HDP framework.

Teh *et al.*[13,14] give a brief introduction to applying HDP for the iHMM. Here we perform a more expansive study of information sharing between multiple HMMs using the iHMM, and provide a detailed inference algorithm. We show

results and analysis for continuous HMMs, and give examples of a new application, in particular multiaspect target detection using the iHMM. We demonstrate how the iHMM may be used to build a single model for a class of targets, and how the learned model yields information about the physical characteristics and the relationships between the targets within the class. These ideas are demonstrated using measured acoustic scattering data, with comparison to more traditional techniques.

The remainder of the paper is organized as follows. In Sec. II we briefly review the basic technical definition of DPs and discuss the DP mixture model. Section III introduces the HDP and provides its interpretation. With these representations as background, we describe in Sec. IV the iHMM and the corresponding Markov chain Monte Carlo (MCMC) algorithm for inference. We present experimental results on measured underwater scattering data in Sec. V. Conclusions are summarized in Sec. VI.

II. DIRICHLET PROCESSES

Let H be a measure on a space Θ, and let γ be a positive real number. A DP[10] is a distribution for a random density function $G(\theta)$, denoted by $G \sim DP(\gamma, H)$; while most distributions are defined on random variables, we emphasize that $DP(\gamma, H)$ is a distribution on the distribution G, the latter defined for an associated set of random variables θ. The "base" distribution H provides the prior information of G with $E[G] = G_0$, and the concentration parameter γ controls how similar G is to H.

The density function G drawn from $DP(\gamma, H)$ is discrete with probability one,[10] a property made explicit by the stick-breaking construction[16]

$$G(\theta) = \sum_{k=1}^{\infty} \beta_k \delta_{\alpha_k^*}, \tag{1}$$

where $\delta_{\theta_k^*}$ is a discrete measure concentrated at θ_k^*. The countably infinite random parameters $\{\theta_k^*\}_{k=1}^{\infty}$ are independently sampled from the base measure H, which is assumed nonatomic [i.e., $H(\theta^*) = 0$ for any θ^*]. Once the set of θ^*'s (the support of G) is instantiated, a θ drawn from G takes value θ_k^* with probability β_k. The weights $\{\beta_k\}_{k=1}^{\infty}$ are defined by a beta distribution that partitions a unit-length "stick" as follows:

$$\beta_k = \beta_k' \prod_{l=1}^{k-1} (1 - \beta_l') \quad \beta_k' \sim \text{beta}(1, \gamma) \quad \text{for } k = 1, 2, \dots . \tag{2}$$

The discreteness of G implies that samples drawn from G have a clustering property. Let $(\theta_1, \theta_2, \dots, \theta_n)$ be independent and identically distributed (i.i.d) random variables sampled from G. If we integrate over G, we have the conditional distribution of θ_i as
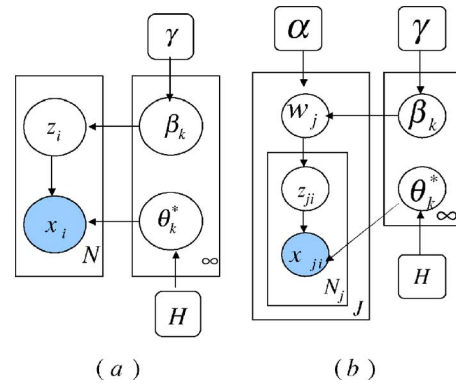
FIG. 1. (a) Graphical representation of a DP mixture model. Circles denote random variables and arrows denote the dependence. Rounded squares represent fixed hyperparameters and the rectangle denotes replications. The bright circle is the observed data. (b) Graphical representation of a hierarchical Dirichlet process (HDP) mixture model.

$$\theta_i | \boldsymbol{\theta}^{-i}, \gamma, H \sim \sum_{k=1}^{M} \frac{n_k}{i - 1 + \gamma} \delta_{\theta_k^*} + \frac{\gamma}{i - 1 + \gamma} H, \tag{3}$$

where $\theta_1^*, \dots, \theta_M^*$ denote the distinct values (clusters) taken by the set $\boldsymbol{\theta}^{-i} = \{\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n\}$ and n_k is the number of θ's equal to θ_k^*.[17] Equation (3) indicates that the DP framework encourages parameter sharing, and the number of clusters is flexible and inferred from the data (through the likelihood function used to yield the posterior for the parameters[11]). The hyperparameter γ controls how often a new cluster is created. In the limit as $\gamma \to \infty$, each θ_i is a cluster and drawn independently from H, resulting in $G = H$.

A Dirichlet process is commonly used as a nonparametric prior distribution for a mixture model with unbounded number of components.[11] Assume the observation x_i is generated from a distribution $F(\theta)$ with parameter θ_i. The density function on the θ's is G, which is assumed to be drawn from $DP(\gamma, H)$, and thus we have a Dirichlet process mixture model. A graphical representation of a DP mixture model is given in Fig. 1(a), where G is represented by the stick-breaking construction (1). Indicator variable z_i denotes the mixture component generating the data point $x_i \sim F(\theta_{z_i}^*)$, i.e., $\theta_i = \theta_{z_i}^*$. Unlike standard mixture models with which we have to set a fixed initial number of mixture components, with DP as a prior on G the number of mixture components is treated as infinite and the actual (finite) components used by the mixture model are inferred automatically from the data.[17] For example, if each mixture component $F(\theta)$ is a Gaussian distribution with θ representing the mean and variance, using DP as a prior yields the infinite Gaussian mixture model.[18]

III. HIERARCHICAL DIRICHLET PROCESSES

Teh *et al.*[13,14] proposed a nonparametric Bayesian approach, namely, the hierarchical Dirichlet process (HDP) to describe a set of coupled mixture models. Assume we have J groups of data. To construct an HDP, a global probability measure G_0 is first drawn from a Dirichlet process $G_0 \sim DP(\gamma, H)$, and then G_j's are drawn independently from $DP(\alpha, G_0)$ for each group:

$$G_j | \alpha, G_0 \sim DP(\alpha, G_0),$$

Ni *et al.*: Infinite hidden Markov model

$$G_0 | \gamma, H \sim \mathrm{DP}(\gamma, H). \qquad (4)$$

The discreteness of G_0 [as shown in Eq. (1)] guarantees that the G_j's will reuse the same set of shared mixture components defined in G_0 but with different proportions:[13]

$$G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k^*}, \quad \{\theta_k^*\}_{k=1}^{\infty} \sim H,$$

$$G_j = \sum_{k=1}^{\infty} w_{jk} \delta_{\theta_k^*} \quad \mathbf{w}_j \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}), \qquad (5)$$

where $\boldsymbol{\beta} = \{\beta_1, \beta_2, \dots\}$ is an infinite-dimensional probability mass function (pmf) defined in Eq. (2) and $\mathbf{w}_j = \{w_{j1}, w_{j2}, \dots\}$ is also an infinite-dimensional pmf.

The graphical model of an HDP mixture is shown in Fig. 1(b). Assume θ_{ji} is the parameter associated with datum x_{ji} in group j, where each group has N_j data samples. Similar to the DP mixture model, let $\{\theta_k^*\}_{k=1}^{\infty}$ be the set of shared mixture components defined in G_0 and z_{ji} is the indicator variable for which $\theta_{ji} = \theta_{z_{ji}}^*$, and each data point is generated independently from $x_{ji} \sim F(\theta_{z_{ji}}^*)$. According to Eq. (5), the HDP mixture model can be written as

$$x_{ji} | z_{ji}, \quad \{\theta_k^*\}_{k=1}^{\infty} \sim F(\theta_{z_{ji}}^*),$$

$$\{\theta_k^*\}_{k=1}^{\infty} | H \sim H,$$

$$z_{ji} | \mathbf{w}_j \sim \mathrm{Mult}(\mathbf{w}_j),$$

$$\{\mathbf{w}_j\}_{j=1}^{J} | \alpha, \boldsymbol{\beta} \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}),$$

$$\boldsymbol{\beta} | \gamma \sim \mathrm{Stick}(\gamma), \qquad (6)$$

where $\mathrm{Stick}(\gamma)$ are the weights from the stick-breaking process with parameter γ and $\mathrm{Mult}(\mathbf{w}_j)$ is a multinomial distribution with parameter $\mathbf{w}_j$.

The generative process underlying HDP can be understood as the Chinese restaurant franchise (CRF).[13] In this setup, there is a globally shared menu G_0 across J restaurants. The customers x_{ji} (associated with θ_{ji}) in restaurant j sitting at table t share a single dish θ_{jt}^*. G_j defines the jth restaurant sharing structure defined by the customer sitting plan. Integrating out the G_j using Eq. (3), we have

$$\theta_{ji} | \boldsymbol{\theta}_j^{-i}, \alpha, G_0 \sim \sum_{t=1}^{T_j} \frac{n_{jt}}{N_j - 1 + \alpha} \delta_{\theta_{jt}^*} + \frac{\alpha}{N_j - 1 + \alpha} G_0, \qquad (7)$$

where n_{jt} is the number of customers in restaurant j sitting at table t and sharing parameter (dish) θ_{jt}^* and $\boldsymbol{\theta}_j^{-i} = \{\theta_{j1}, \dots, \theta_{ji-1}, \theta_{ji+1}, \dots, \theta_{jN_j}\}$. The globally shared menu G_0 plays the role of the distribution over dishes θ_{jt}^* across the franchise. If we integrate out G_0, we have the sharing structure among the restaurants, which links all the dishes in the franchise
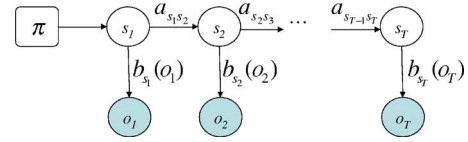
FIG. 2. Graphical representation of an HMM.

$$\theta_{jt}^* | \boldsymbol{\Theta}_1^*, \dots, \boldsymbol{\Theta}_{j-1}^*, \boldsymbol{\Theta}_j^{*(-\mathbf{t})}, \gamma, H \sim \sum_{k=1}^{M} \frac{m_k}{\sum_k m_k + \gamma} \delta_{\theta_k^*}$$

$$+ \frac{\gamma}{\sum_k m_k + \gamma} H, \qquad (8)$$

where $\boldsymbol{\Theta}_r^* = \{\theta_{r1}^*, \dots, \theta_{rT_r}^*\}$ indicates all the dishes in a particular restaurant r, $\boldsymbol{\Theta}_j^{*(-t)}$ indicates all the dishes in restaurant j except the dish θ_{jt}^*, m_k is the number of restaurant dishes (double subscriptions) equal to θ_k^* in *all* restaurants, and M denotes the number of existing kinds of global dishes across the franchise. In the CRF, customers prefer sitting at table t serving dish θ_{jt}^* where many other people n_{jt} are already seated, as in Eq. (7), but sometimes choose a new table. When a new table is created, popular kinds of dishes already served among the franchise are more likely to be ordered, while a new kind of dish (not used by any restaurant thus far) can also be introduced from H, represented by the right-most term in Eq. (8). The dishes (mixture components) in HDP are shared not only within a group, but also across different groups.

IV. INFINITE HIDDEN MARKOV MODEL

A. Hidden Markov models

The idea of using HDP to describe coupled mixture models with unbounded components can be applied to solve fundamental problems in more complex models, such as the HMM. Hidden Markov models have been used widely for modeling sequential data, such as in speech recognition[19] and target classification.[1,2,20] An HMM includes a sequence of observations $\boldsymbol{o} = \{o_1, o_2, \dots, o_T\}$ and a sequence of hidden underlying states $\boldsymbol{s} = \{s_1, s_2, \dots, s_T\}$ which follow a Markov process. In many applications, the first-order Markov chain is used, i.e., the current state variable only depends on the previous state regardless of the history $P(s_t | s_1, s_2, \dots, s_{t-1}) = P(s_t | s_{t-1})$; the observation is then independently generated given the state.

Typically maximum-likelihood HMM parameter learning is performed via the EM algorithm,[19] with the number of states assumed known. Assume we use an HMM with M states and C possible observations (later we generalize to the continuous HMM, for which the observations are not discretized). The parameters of the model are $(\pi^M, A^{M \times M}, B^{M \times C})$, with π being the initial-state probability, A the transition matrix of $P(s_t | s_{t-1})$, and B the observation matrix of $P(o_t | s_t)$. Figure 2 shows the HMM model.

An M-state HMM can be considered as a set of M coupled mixture models, where the mixture components are shared among all the models. Given the hidden state (random

variable) $s_{t-1}=j$, we have the joint distribution of s_t and o_t as $p(o_t,s_t=i|s_{t-1}=j)=a_{ji}b_i(o_t)$, where a_{ji} is the ith element of jth row in A, and $b_i(o_t)$ is the ith row distribution of B. Summing over s_t we have

$$p(o_t|s_{t-1}=j) = \sum_{i=1}^{M} a_{ji}b_i(o_t). \tag{9}$$

This is a finite mixture model where the row of A is the mixture weights and B defines the mixture components. In an M-state HMM there are M such mixture models. Each row of A defines distinct mixture weights for each model and all the models share the same set of mixture components (M components) defined in B. For datum o_t, the random variable s_{t-1} selects the mixture model to be used and the random variable s_t selects the component from which o_t is generated. The form of $b_i(\cdot)$ varies depending on the properties of the data: If the data are discretized, we use a multinomial distribution for $b_i(\cdot)$; and if continuous features are adopted, we use a Gaussian distribution for each $b_i(\cdot)$ and the continuous HMM may be viewed as M coupled Gaussian mixture models.

B. HMM for multiaspect target sensing

Using HMMs for multiaspect target detection and classification has been described in Refs. 1, 2, and 4. Here we provide a brief summary. We use this discussion to motivate the need for the iHMM, explaining for example that the number of underlying HMM states is typically unknown *a priori*. Moreover, the HMM is only a model for the target scattering, and therefore there may not be a single "correct" number of states, motivating the need for inferring a full posterior for the underlying number of states.

The data collected when sensing a target from T target-sensor orientations are represented as a sequence of T scattered wave forms. Although the scattered wave forms are strongly dependent on the (hidden) target-sensor orientation, there is a set of "states" that represent a set of angles over which the scattering physics is relatively stationary. Within each state, the signal statistics characterize the variability in the scattering physics at the corresponding set of target-sensor orientations.

To a good approximation the transitions from one state to another on consecutive measurements can be modeled as a Markov process. Moreover, the target is usually distant or concealed and the actual target-sensor orientation (sampled state) is unknown. This motivates an HMM for representation of each target. The wave forms are typically represented in terms of sets of features, which are here extracted using physics-based matching-pursuits.[1–3]

For target detection and classification, assume target i is assigned a distinct HMM$_i$, trained on the labeled sequential data associated with the target. Once all the target-dependent HMMs are built, the likelihood that testing data $\boldsymbol{o}=\{o_1,o_2,\ldots,o_T\}$ is associated with target i is
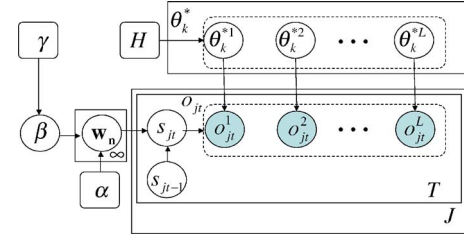
FIG. 3. The infinite hidden Markov model interpreted as an HDP. For observation o_{jt}, s_{jt-1} defines the mixture model to be used and s_{jt} selects the mixture component according to infinite dimension weight vector $w_{s_{jt-1}}$.

$$p(\boldsymbol{o}|\theta_i) = \sum_{s} \pi_{s_1} \cdot \prod_{t=2}^{T} a_{s_{t-1}s_t} \cdot \prod_{t=1}^{T} b_{s_t}(o_t), \tag{10}$$

where the a's and b's are elements of the matrices A and B. The testing data sequences are deemed to be associated with the training target for which the corresponding HMM yields the highest likelihood (i.e., we assume a prior that states that all targets are equally probable). If the likelihoods of all HMMs are below a prescribed threshold, the unknown object is classified as clutter, i.e., as none of the targets associated with the testing data.

The traditional method for estimating HMM parameters is the expectation maximization (EM) method, specified by the Baum-Welch algorithm.[19] However, in this estimation procedure the model structure must be specified in advance, i.e., the number of states is fixed. Rather than performing model selection to select a fixed number of states,[21] we employ a fully Bayesian approach in which the number of states is not fixed *a priori*. The nonparametric HDP is employed to implicitly construct an HMM with an infinite number of states (iHMM), with the posterior on the number of states usually peaked about a finite number of states characteristic of the sequential data used for model training.

C. iHMM via HDP

As described in Sec. IV A, an M-state HMM can be considered as a set of M coupled finite mixture models (each with M mixture components). The value of state s_{t-1} indexes a specific row of the transition matrix serving as the mixture weights for choosing current state s_t. Given the value of s_t, the observation o_t is sampled from the observation model (mixture component) indexed by s_t. To consider an infinite number of states, it is natural to use a set of state-specific DPs, one for each value of the state. Furthermore, these DPs must be shared because they use the same set of mixture components defined in the observation matrix. This is similar to the HDP mixture model but with a key difference being that the data (observations) to group is fixed in the HDP while the group partition is random (indexed by the hidden previous state) in the iHMM.

The HDP construction of the iHMM for multiaspect target sensing is shown in Fig. 3, with parameters defined as

$$o_{jt}|s_{jt}, \quad \{\theta_k^*\}_{k=1}^{\theta} \sim F(\theta_{s_{jt}}^*),$$

$$\{\theta_k^*\}_{k=1}^{\infty}|H \sim H,$$

$$s_{jt}|s_{jt-1}, \quad \{\mathbf{w}_n\}_{n=1}^\infty \sim \text{Mult}(\mathbf{w}_{s_{jt-1}}),$$

$$\{\mathbf{w}_n\}_{n=1}^\infty|\alpha,\boldsymbol{\beta} \sim \text{DP}(\alpha,\boldsymbol{\beta}),$$

$$\boldsymbol{\beta}|\gamma \sim \text{Stick}(\gamma), \tag{11}$$

where $\mathbf{w}_n$ corresponds to the row of transition matrix A, and $F(\theta_k^*)$ corresponds to the observation model $b_k(\cdot)$. Each sampled scattered wave form (observation) is represented with an L-dimensional feature vector $o_{jt}=[o_{jt}^1,\ldots,o_{jt}^L]$. Each feature vector is assumed to be generated from a model F. When considering a continuous HMM, F is an L-dimensional Gaussian and θ represents the mean vector and covariance matrix. As discussed further in the following, the base distribution H is selected to be conjugate to F for better computational speed. For a discrete HMM θ represents the parameters of a multinomial distribution.

Considering the iHMM from the Chinese restaurant franchise view of the HDP, without loss of generality, we write the sequence of observations as $\{o_1,\ldots,o_T\}$ and the corresponding hidden state sequence as $\{s_1,\ldots,s_T\}$; the customer is o_t and o_t enters the restaurant indexed by the state s_{t-1}. The customer chooses table s_t according to $A(s_{t-1},:)$, the transition probability from s_{t-1}. If the customer sits at an existing table, she eats the same dish $b_{s_t}(\cdot)$ on that table; the parameters associated with $b_{s_t}(\cdot)$ are the mean and covariance for continuous observations, and a probability mass function for discrete observations. If a new table is chosen, a dish is selected according to its popularity among all the restaurants and sometimes a new kind of dish is chosen from H. Note that for customer o_t, the previous state s_{t-1} is the restaurant (group) index and the current state s_t is the table (component) index in that restaurant, while for customer o_{t+1}, the s_t becomes a restaurant (group) index. Thus the group index and the component index are coupled. This is a key difference between the HDP and the iHMM. Furthermore, the group assignment is decided by the hidden state, which is from a countably infinite set, thus there are also a countably infinite number of groups. A comparison of the HDP and the iHMM is given in Fig. 4.

D. Inference and learning

We utilize a MCMC sampling scheme for the iHMM. Specifically, the iHMM learning is based on an extension of the HDP Gibbs sampler[13,14] with temporal information considered.

	HDP (CRF view)	iHMM
Group	Restaurant j (fixed)	s_{t-1} (random)
Data	Customer x_{ji}	o_t
Component	Dish $\theta_{ji}^*=\theta_k^*$, $k=1\sim\infty$, $\theta_k\sim H$	$b_{s_t}(\cdot)$, $s_t=1\sim\infty$
DP weights	Popularity $\mathbf{w}_j$	$A(s_{t-1},:)$
Likelihood	$F(x_{ji}|\theta_{ji}^*=\theta_k^*)$	$b_{s_t}(o_t)$

FIG. 4. Comparison of grouped data structure under the HDP and the iHMM.

Assume we have a sequence of observations $\boldsymbol{o} = \{o_1,\ldots,o_T\}$ associated with a hidden state sequence $\boldsymbol{s} = \{s_1,\ldots,s_T\}$. The group partition of o_t is indexed by the previous hidden state s_{t-1}. Given the initialization of the hidden state sequence $\boldsymbol{s}$, we have the group partition plan of the observations. Let x_{ji} (corresponding to an o_t) denote the ith observation in group j, and x_{ji} is drawn from $F(\theta_{z_{ji}}^*)$, with $z_{ji}=k$ denoting the mixture component associated with x_{ji}. Thus, we have two equivalent views of the data: The sequential data view $(\mathbf{o},\mathbf{s})$ and the group data view $(\mathbf{x},\mathbf{z})$, where the groups are coupled via the underlying state sequence. Further let n_{jk} denote the number of data points in group j sharing the same component k, and m_{jk} denote the number of clusters using component k in group j.

In the Chinese restaurant franchise, the data are first assigned to a table t_{ji}, then the tables are assigned to some mixture component k_{jt}. This indirect association from data to mixture component, viewed by Beal et al.[22] complicates the iHMM because the group partition of the iHMM itself is random. We propose a method for which the mixture components are assigned directly to the data. Considering an instantiation of the global DP G_0, we write the mixture weights as $\boldsymbol{\beta}=\{\beta_1,\ldots,\beta_M,\beta_u\}$, where β_1 through β_M correspond to the M existing mixture components used by the data, and β_u represents all of the infinite components which are not used yet. The variables of interest in this sampling scheme are $\boldsymbol{\beta}$, z, and $\boldsymbol{m}$ mentioned earlier. The hyperparameters are α and γ that are assumed to be fixed at this point.

1. Gibbs sampling of random variables

a. Sampling z. The conditional distribution of the index variable z_{ji} given the remaining variables can be derived from Eqs. (7) and (8). Let s_t be the hidden state in the sequential data view and z_{ji} be the equivalent variable in the group data view. Let s_{t+1}, associated with some other z, be the next state value, then z_{ji} can be sampled as follows:

$$p(z_{ji}=k|\mathbf{z}^{-ji},\mathbf{m},\boldsymbol{\beta}) \propto \begin{cases} (n_{jk}^{-ji}+\alpha\beta_k)f_k^{-x_{ji}}(x_{ji})\dfrac{\alpha\beta_{s_{t+1}}+n_{ks_{t+1}}}{a+\displaystyle\sum_{k'=1}^M n_{kk'}} & k=1,\ldots,M \\[2em] \alpha\beta_u f_{k^{\text{new}}}^{-x_{ji}}(x_{ji})\beta_{s_{t+1}}, & k=k^{\text{new}}, \end{cases} \tag{12}$$

where the bold variable represents the collection of that kind of variables and the superscript refers to the element left out from the set of variables. For example, $z = (z_{ji}: \text{all } j, i)$ is the set of all the index variables and z^{-ji} means the element z_{ji} is removed from the index variables set z.

The posterior distribution of z_{ji} taking value k consists of three parts: Incoming probability, observation likelihood, and outgoing probability.

The incoming probability $(n_{jk}^{-ji} + \alpha\beta_k)$ is derived from an HDP prior. The first term in the sum is the number of customers eating dish k in restaurant j, which comes from a single-restaurant DP, and the second term accounts for choosing a new table. A new table is chosen with a probability proportional to α and with probability β_k the dish served on new table is the same as a previously used dish k, and with probability β_u a totally new dish k^{new} is chosen. Note that when a new component (dish) is sampled, we draw $a \sim \text{beta}(1, \gamma)$ and set $\beta_k^{\text{new}} = a\beta_u$ and $\beta_u^{\text{new}} = (1-a)\beta_u$. This procedure can be understood as follows: Once a new component is instantiated, it breaks the remaining stick with probability a. Using the fact that the sequence of stick-breaking weights is a length-biased permutation of the weights, as shown in Ref. 23, the weight a will have the same distribution as any other stick-breaking weight, i.e., $\text{beta}(1, \gamma)$.

The data likelihood component refers to evaluation of the observation x_{ji} given the parameter $\theta_{z_{ji}=k}^*$. Assuming the base measure H is conjugate to data distribution F, we can analytically integrate out the mixture component parameters θ_k^*. Let $f(\cdot|\theta)$ and $h(\cdot)$ denote the density function of F and H, respectively, the data likelihood is calculated as

$$f_k^{-x_{ji}}(x_{ji}) = \frac{\int f(x_{ji}|\theta_k^*) \prod_{j'i' \neq ji, z_{j'i'}=k} f(x_{j'i'}|\theta_k^*) h(\theta_k^*) d\theta_k^*}{\int \prod_{j'i' \neq ji, z_{j'i'}=k} f(x_{j'i'}|\theta_k^*) h(\theta_k^*) d\theta_k^*}.$$

(13)

If a new component k^{new} is selected, the data likelihood is simply

$$f_{k^{\text{new}}}^{-x_{ji}}(x_{ji}) = \int f(x_{ji}|\theta_k^*) h(\theta_k^*) d\theta_k^*.$$

(14)

The outgoing probability component arises from the sequential property in the Markov chain. Since we assume a first-order Markov chain in our HMM, the conditional distribution of s_t (equivalent to z_{ji}) given all the other states is only dependent on the previous state s_{t-1} and the next state s_{t+1}. If the state s_t takes value k, then the outgoing probability is proportional to the transitions that already happened from k to s_{t+1} in group k, counted as $n_{ks_{t+1}}$, as well as the transitions counted from selecting a new table, given by $\alpha\beta_{s_{t+1}}$. In addition, the counting must be normalized within each group as different groups have different total outgoing transitions. If s_t is the last state of the sequence, then a uniform distribution is adopted for the outgoing probability.

If, as a result of updating z_{ji}, some component k becomes unallocated, i.e., $n_{jk}=0$ for all j, then the corresponding β_k will become zero and the probability that this component will be used again in the future is zero, since that is always proportional to $n_{jk}+\beta_k$. Therefore we may delete this mixture component as well as adjust the global weight vector $\boldsymbol{\beta}$.

b. Sampling **m**. The tables in this sampling scheme are represented in terms of m_{jk}, denoting the number of tables serving dish k in restaurant j. Given all the other variables, the distribution of m_{jk} is a DP with concentration parameter $\alpha\beta_k$ and number of customers n_{jk}. Given n_{jk} customers in restaurant j eating dish k, m_{jk} may be understood as the number of tables occupied by those customers. Recall that a customer x_{ji} will sit at an existing table serving dish k with probability proportional to n_{jk}^{-ji} and choose a new table along with dish k with a probability proportional to $\alpha\beta_k$, thus the number of tables m_{jk} can be regarded as the number of tables in a special balcony which only holds the customers who are eating dish k in restaurant j. This turns out to be a DP with parameter $\alpha\beta_k$ and n_{jk} customers. The exact conditional distribution over m_{jk} has been shown by Antoniak[24]

$$p(m_{jk} = m|\mathbf{z}, \mathbf{m}^{-jk}, \boldsymbol{\beta}) = \frac{\Gamma(\alpha\beta_k)}{\Gamma(\alpha\beta_k + n_{jk})} s(n_{jk}, m)(\alpha\beta_k)^m,$$

(15)

where $s(n, m)$ are unsigned Stirling numbers of the first kind. By definition $s(0,0) = s(1,1) = 1$, $s(n,0) = 0$ for $n > 0$ and $s(n,m) = 0$ for $m > n$. Other entries can be computed as $s(n+1, m) = s(n, m-1) + ns(n, m)$.

Sampling for β simply follows the definition of the DP weights

$$(\beta_1, \ldots, \beta_M, \beta_u)|\mathbf{z}, \mathbf{m} \sim \text{Dir}\left(\sum_j m_{j1}, \ldots, \sum_j m_{jM}, \gamma\right).$$

(16)

2. Hyperparameters sampling

In the above section we assumed the hyperparameters were fixed for sampling the random variables of interest. Here we relax this assumption by sampling the posterior distribution of hyperparameters given $\mathbf{z}$, $\mathbf{m}$, and $\boldsymbol{\beta}$.

For a given group, the index variables $z_j = \{z_{j1}, \ldots, z_{jn_j}\}$ are multinomial distributed with parameter $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_M\}$ associated with that group,

$$p(z_{j1}, \ldots, z_{jn_j}|\boldsymbol{\pi}) = \prod_{k=1}^{M} \pi_k^{n_{jk}},$$

(17)

with n_{jk} denoting the number of data points under mixture component k and n_j the total number of data in group j. Since the group proportion $\boldsymbol{\pi}$ is a DP with concentration parameter $\alpha\boldsymbol{\beta}$,

$$p(\boldsymbol{\pi}|\alpha, \boldsymbol{\beta}) \sim \text{Dir}(\alpha\beta_1, \ldots, \alpha\beta_M) = \frac{\Gamma(\alpha)}{\prod_{k=1}^{M} \Gamma(\alpha\beta_k)} \prod_{k=1}^{M} \pi_k^{\alpha\beta_k - 1},$$

(18)

we can analytically integrate out $\boldsymbol{\pi}$ to yield
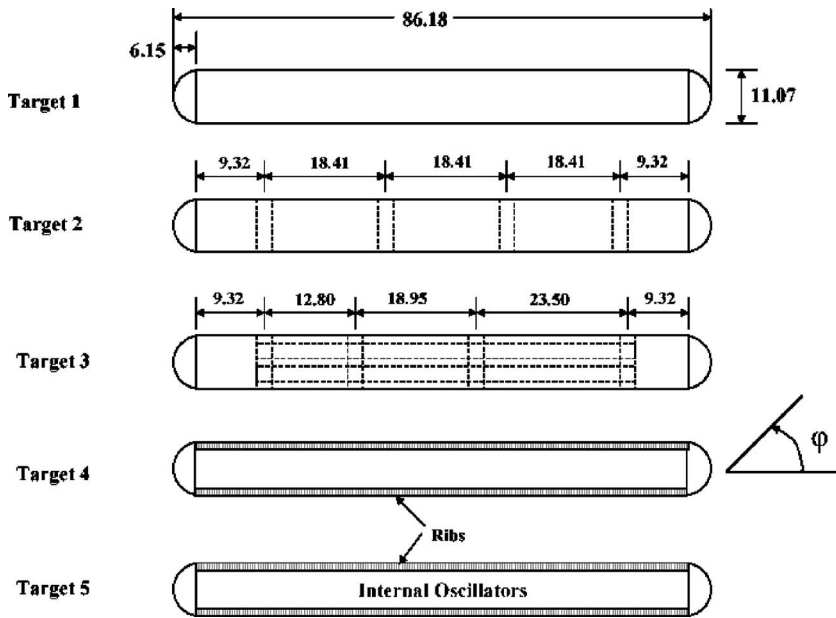
FIG. 5. Five elastic shell targets, with all units in centimeters. The scattered fields are observed as a function of angle at a fixed distance from the target center, in a plane bisecting the target axis. The nominal external sizes and shapes of the targets are the same, with the acoustic scattering distinguished by the internal structure.

$$p(z_j | \alpha, \boldsymbol{\beta}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + n_j)} \prod_{k=1}^{M} \frac{\Gamma(\alpha \beta_k + n_{jk})}{\Gamma(\alpha \beta_k)}. \tag{19}$$

Here we place a vague gamma prior on the hyperparameter $\alpha \sim \mathcal{G}(a, b) = b^a / \Gamma(a) \alpha^{a-1} e^{-b\alpha}$ and the posterior can be derived from Eq. (19)

$$p(\alpha | \mathbf{z}, \mathbf{m}, \boldsymbol{\beta}) \propto \mathcal{G}(a_\alpha, b_\alpha) \prod_{j=1}^{M} \frac{\Gamma(\alpha)}{\Gamma(\alpha + n_j)} \prod_{k=1}^{M} \frac{\Gamma(\alpha \beta_k + n_{jk})}{\Gamma(\alpha \beta_k)}. \tag{20}$$

Similarly we obtain the posterior of hyperparameter γ with

$$p(\gamma | \mathbf{z}, \mathbf{m}, \boldsymbol{\beta}) \propto \mathcal{G}(a_\gamma, b_\gamma) \frac{\Gamma(\gamma)}{\Gamma\left(\gamma + \sum_{jk} m_{jk}\right)}$$

$$\times \prod_{k=1}^{M} \frac{\Gamma\left(\gamma \beta_k + \sum_{j} m_{jk}\right)}{\Gamma(\gamma \beta_k)}. \tag{21}$$

Note that γ controls the expected number of mixture components used among all the groups and α governs the distribution over the number of components used in each group independently.
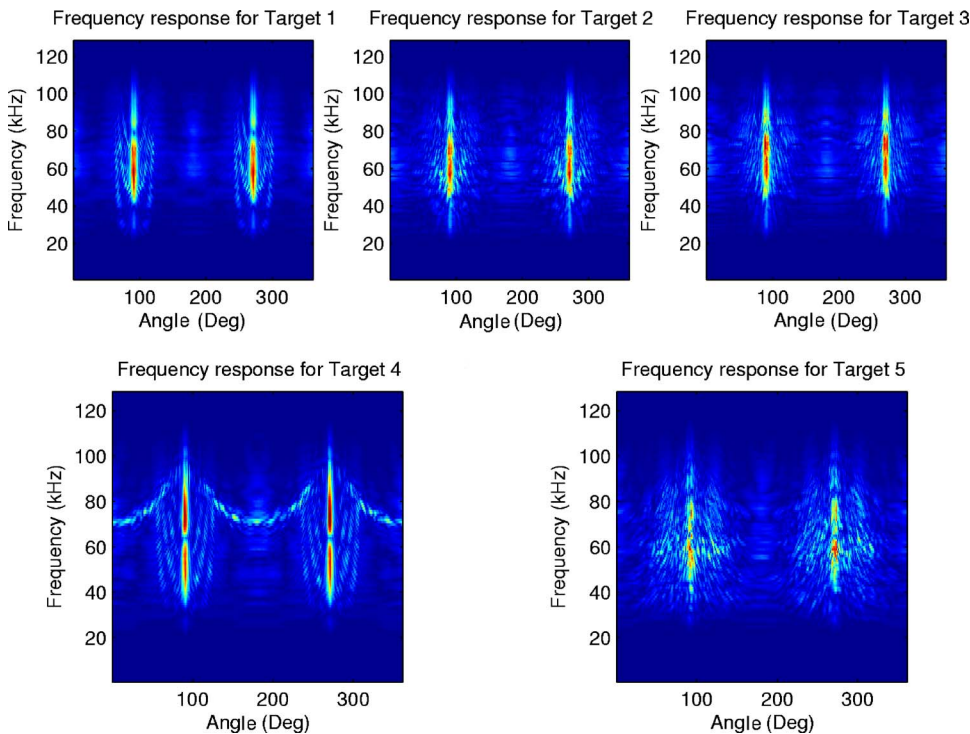
FIG. 6. Scattering data of five elastic targets as shown in Fig. 5.
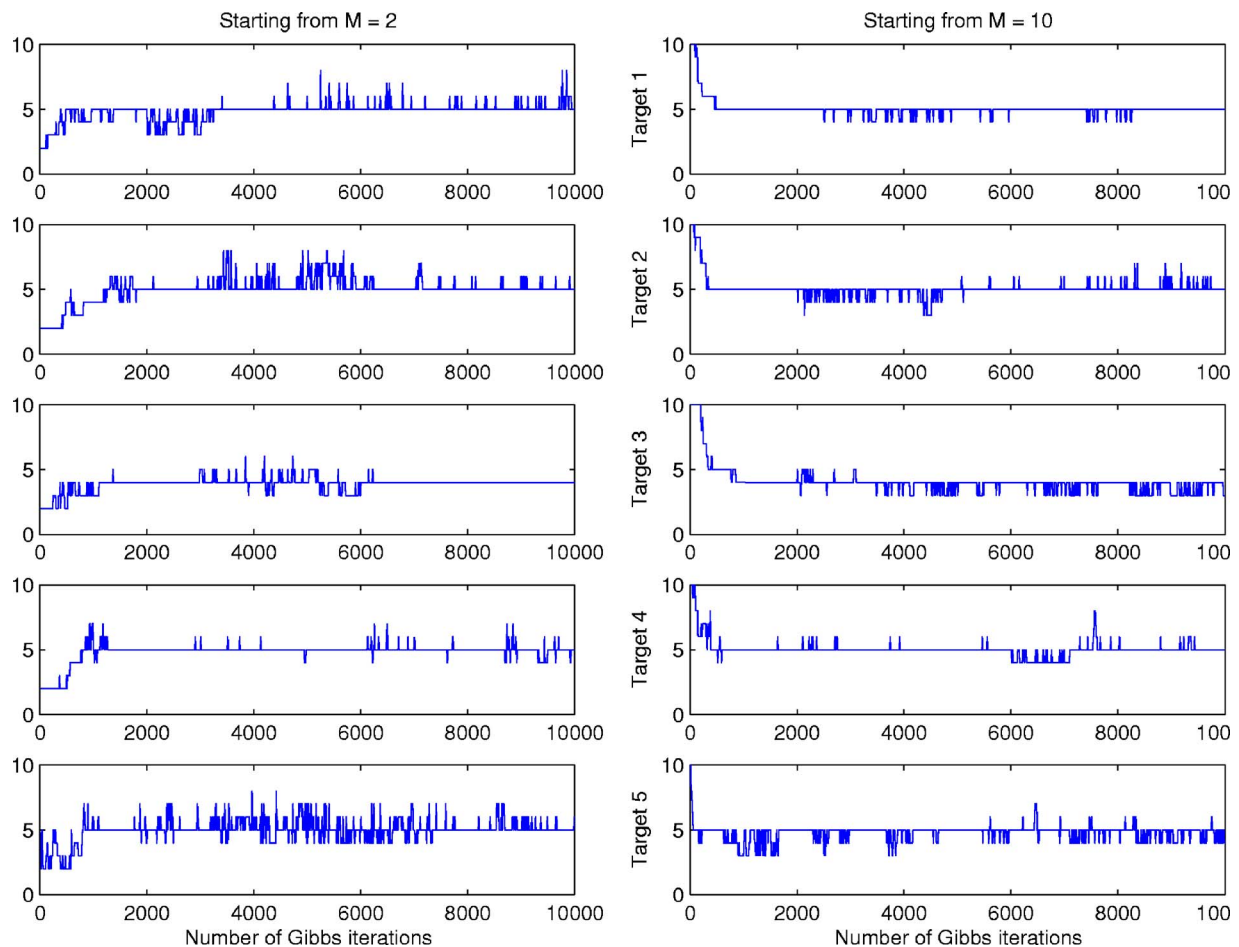
FIG. 7. Trajectories of number of underlying states on Shell data shown in Fig. 6. Each column is associated with the studying case with different initial number of states. The results for each target are shown in one row.

E. Testing likelihoods

At any iteration of Gibbs sampling, we can build an HMM based on the posterior sample of the hidden state sequence $\mathbf{s}$ (equivalent to $\mathbf{z}$). The transition matrix A is obtained by counting the number of transitions given the sampled values of the hidden state sequence, and the observation model is built on all the observed data generated from a particular state. To compute the initialization state probabilities $\boldsymbol{\pi}$; we create a virtual state $s_0 = 0$ with transitions only to s_1 for each of the sequences, therefore $\boldsymbol{\pi}$ is computed from the transitions in the special group 0. In the iHMM, a set of P independent posterior samples (associated with P HMMs) are obtained to represent the posterior distribution over the hidden state sequence, and the testing likelihood for any testing sequence is the average likelihood of the P models.

V. EXPERIMENT RESULTS

A. Modeling for acoustic-scattering from real targets

We consider multiaspect acoustic scattering data from five elastic shell targets. Details on the targets, scattering data, and features are described in Refs. 1 and 2. We give a brief summary here for completeness. The physical description of the five targets is shown in Fig. 5. The scattered fields are observed as a function of angle with data collected in 1° increments. Features for each target are then extracted from the time-domain scattered fields using matching pursuits.[1–3] The feature vectors across all the target-sensor orientations and all types of targets are concatenated to perform vector quantization (VQ).[25] Therefore the observation is a discrete element mapped from the feature vector to a code within a VQ codebook of size 20.

The scattering data of the five targets are shown in Fig. 6. We plot the magnitude of the frequency response as a function of sensing angle (over 360° target-sensor orientations) for each target. In previous work,[1,2] five distinct states were defined manually over a segment of 90°, and the remaining states at other angles are manifested using the symmetry of the target reflected in Fig. 5. Here we apply the iHMM to discover the number of states from the training data, without setting the model structure in advance. Note that for this real data the HMM is simply a model, and therefore there may not be a clear (single) number of states for the observed data. Importantly, the iHMM will provide an estimate of the full posterior on the number of states for each target.

In the training process we use a small number of angular samples from the data. Each data sequence consists of five observations that are sampled in 5° increments. We use 114 data sequences from each target for training the model. Figure 7 shows the state evolution trajectories using the iHMM
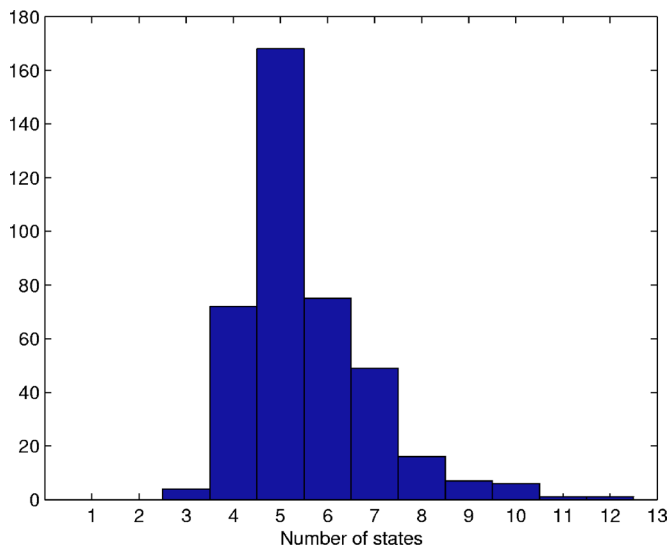
FIG. 8. Posterior distribution on the number of states for target 5 (initialization with $M=2$).

on the five shell targets. We consider iHMM initializations with $M=2$ and $M=10$ states and 10 000 Gibbs iterations are performed for each case.

From Fig. 7 we observe that the iHMM represents each target with approximately five states, except the third target which is modeled with four hidden states. It is interesting to note that the results obtained from the iHMM (number of states) are similar to the results in Ref. 2. In that work, the model structure is defined manually based upon the physical statistics of the elastic target and each target is represented by five distinct states and the resulting models have been successfully applied in target classification. However, the real targets may not be modeled by a single HMM, and fixing the model structure leads to under- or over-fitting. Instead, the iHMM gives a full posterior distribution on the number of states. An example for target 5 (initialized with

$M=2$) is shown in Fig. 8, running 100 000 Gibbs iterations after a 5000 burn-in, collecting 400 posterior samples with a spacing of 250 samples. This posterior distribution gives the information of the proper number of states and eliminates the problem of model selection.

B. Information sharing among multiple targets

The iHMM is able to share data from different HMMs to learn the cumulative set of HMM-dependent states (across all the HMMs, with an appropriate sharing of data between similar states of different HMMs). By using the data from multiple targets (usually each target is modeled by an HMM) collectively, the training data for each target are strengthened and the generalized classification results are improved. In the next experiments, we consider a target detection problem where five shell targets are considered as "targets of interest," and six clutter items are considered as false targets. The six false targets are a small missile-like object, a plastic container, a 55-gallon drum, two rocks, and a log, respectively, with associated data shown in Fig. 9. The details of the shell targets and the clutter are described in Refs. 1, 2, and 20. The goal is to test if a new target is in the family of shell targets or not. We train on four of the shell targets and the item under test is the remaining one shell target (label 1) or one of the six false targets (label 0). All six of the false targets and the held-out shell target are not seen when training. When testing and evaluating performance, we consider each possible test item, as viewed from all possible initial angles (although only a small subset of angles are considered for a particular testing example). Both the training sequences and the testing sequences were of length 5, and the angular sampling between consecutive measurements was 5°.

We compare development of single HMMs (SHMM) designed for each of the four training targets as well as one overarching HMM designed for all four training targets (class-based iHMM), and in both cases the iHMM is used for
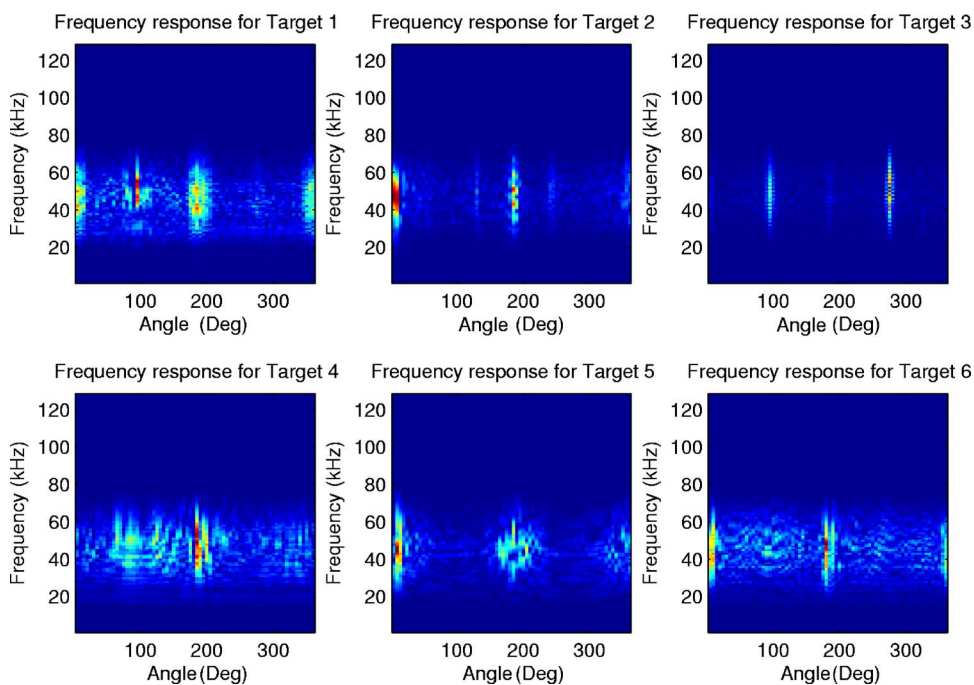
FIG. 9. Scattered fields of six false targets. The six false targets are a small missile-like object, a plastic container, a 55-gallon drum, two rocks, and a log, respectively.
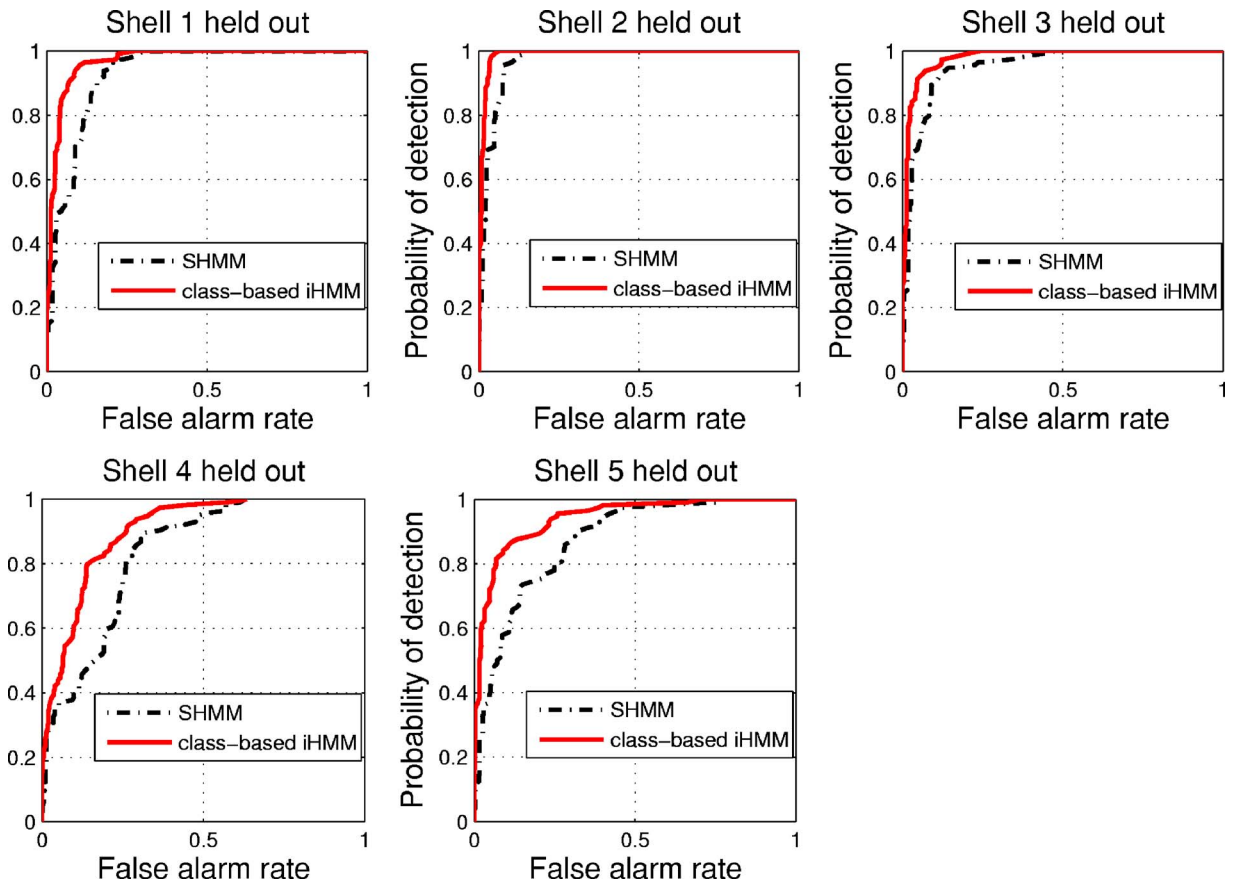
FIG. 10. Detection results for different sets of targets using the SHMM and the class-based iHMM. The results are presented with the respective shell target held-out of the training set.

model design. We use 40 posterior samples (with a spacing of 250 iterations) to compute each HMM likelihood. The testing likelihood for the SHMM is the maximum of the four individual HMM likelihoods. We note that for the SHMM, there are counterpart models that may be designed using the conventional EM algorithm, one model for each shell. The number of HMM states may be inferred for the target-dependent HMMs using physical considerations (note that for the EM solution the number of states is set *a priori*, and fixed). By contrast, for the class-based iHMM (single iHMM for all four training targets), it is difficult to have an analog computed using the EM algorithm, because the proper number of states is difficult to determine (some states associated with the single targets may be shared, and others not shared, and this decomposition is difficult to determine *a priori*). A significant advantage of the iHMM is that it adaptively determines a posterior on the proper number of states, automatically learning the appropriate state sharing structure across the four training targets. We also note that this class-based iHMM, by sharing data across the training targets, will not model any one of the training targets, but will rather infer relationships between the multiple training targets within this class. This generalization is beneficial when testing on targets not seen while training, as long as the testing targets of interest are within the class of training data.

This generalization is reflected in the results presented in Fig. 10, for which the class-based iHMM consistently outperforms the iHMMs designed for each of the training tar-

gets separately. The results are measured by receiver operation characteristic (ROC) curves. We emphasize that the iHMM does not find a single HMM but rather a full posterior distribution on the model structure and we show in Fig. 11 the posterior distribution on the number of states associated with the class-based iHMM (for case 1 in Fig. 10), running 100 000 Gibbs iterations after a 5000 burn-in, collecting 400 posterior samples with a spacing of 250 samples.

The performance gain using the class-based iHMM for cases 2 and 3 are smaller compared with other cases. This is due to the fact that shell targets 2 and 3 possess relatively similar scattered field (Fig. 6). In cases 2 and 3, the SHMM
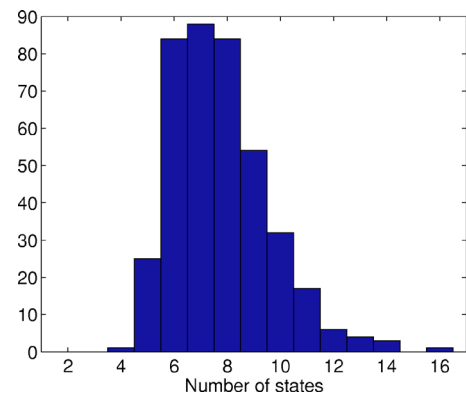
FIG. 11. Posterior distribution on model structure (the number of states) for case 1 in Fig. 10.
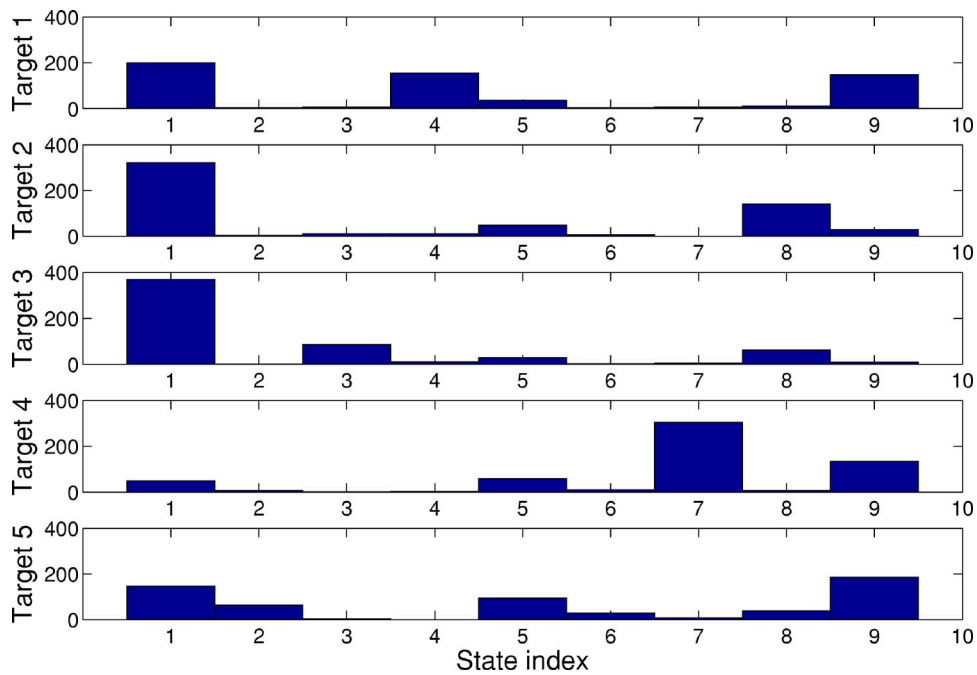
FIG. 12. Hidden state distribution across all five shell targets. Note that many targets within the same class share states.
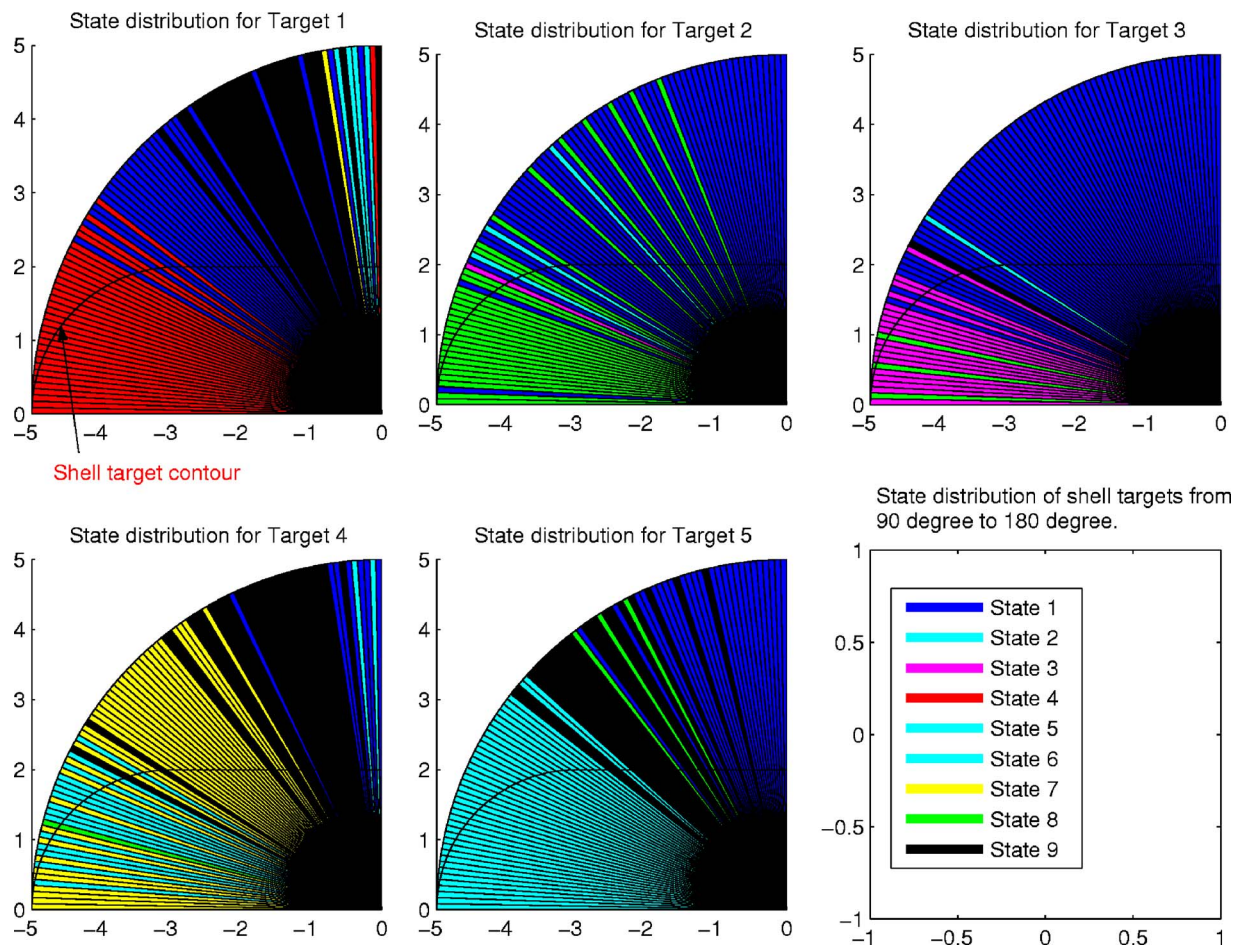
FIG. 13. Relationship between physical angles and HMM states for shell targets. Some angles share the same signal statistics (state) while some are distinct from target to target.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Ni *et al.*: Infinite hidden Markov model   2741

detects the held-out testing shell target well because of having similar targets in the training set and leaving a small margin for improvement using the class-based iHMM. In other cases, the class-based iHMM provides the ability for sharing state information between shell targets and giving better generalization for unseen shell target.

To see the relationship of target-dependent states across the shell targets, we also trained a class-based iHMM on the collection of *all* five shell targets and plot the hidden state distribution based on the last Gibbs iteration, with the result depicted in Fig. 12. We observe that the underlying states are shared over all targets, showing the utility of using the class-based iHMM as compared to the single HMMs.

As mentioned earlier, the HMM "state" corresponds to a sector of target-sensor orientations over which the signal statistics are invariant. Due to the target symmetry presented in Fig. 6, the set of states are uniquely defined in a 90° segment. In Fig. 13, we plot the physical angles in relation to the sampled states from 90° to 180° of a target. It can be seen that at certain angles the targets have similar scattering signal, thus they share the statistics for these angles, while at some other angles, the targets are discriminative. The class-based iHMM explores this sharing structure and strengthens the generative model learning for targets. These types of results have significant potential for learning relationships in the aspect-dependent scattering physics between different targets, solely based on the data.

VI. CONCLUSION

We have considered an extensive study of the iHMM, which provides a principled approach to solve the fundamental problem of model selection in hidden Markov models, here applied to multiaspect acoustic target detection. The iHMM is implemented using a nonparametric tool, namely the hierarchical Dirichlet process (HDP), which infers a full posterior distribution on the number of underlying states from the training data. After discussing the difficulties in extending HDP to iHMM, we have provided a detailed procedure for performing Gibbs sampling for iHMM inference.

Promising iHMM results have been obtained using measured data from real elastic targets. It is shown that the iHMM not only discovers the correct number of hidden states for modeling the data, but also yields the corresponding model parameters in both discrete and continuous HMMs. In addition, our results indicate that the iHMM implicitly shares the underlying state information among different targets, as appropriate. The results on real data demonstrate improved detection performance using the iHMM relative to using other competing methods.

[1] P. Runkle, P. K. Bharadwaj, L. Couchman, and L. Carin, "Hidden Markov models for multi-aspect target classification," IEEE Trans. Signal Process. **47**, 2035–2040 (1990).

[2] P. Runkle, L. Carin, L. Couchman, T. J. Yoder, and J. A. Bucaro, "Multi-aspect identification of submerged elastic targets via wave-based matching pursuits and hidden Markov models," J. Acoust. Soc. Am. **104**, 937–946 (1999).

[3] M. McClure and L. Carin, "Matching pursuits with a wave-based dictionary," IEEE Trans. Signal Process. **45**, 2912–2927 (1997).

[4] P. K. Bharadwaj, P. Runkle, and L. Carin, "Target identification with wave-based matching pursuits and hidden Markov models," IEEE Trans. Antennas Propag. **47**, 1543–1554 (1999).

[5] R. Varriere and R. L. Moses, "High-resolution radar target modeling using a modified prony estimator," IEEE Trans. Antennas Propag. **40**, 13–19 (1992).

[6] L. C. Potter, D. M. Chiang, R. Carriere, and M. J. Gerry, "A gtd-based parametric model for radar scattering," IEEE Trans. Antennas Propag. **43**, 1058–1067 (1995).

[7] M. McClure, R. C. Qiu, and L. Carin, "On the superresolution identification of observables from swept-frequency scattering data," IEEE Trans. Antennas Propag. **45**, 631–641 (1997).

[8] C. E. Baum, *The Singularity Expansion Method* (Springer, New York, 1976).

[9] D. Kralj, L. Mei, T. T. Hsu, and L. Carin, "Short-pulse propagation in a hollow waveguids: Analysis, optoelectronic measurement and signal processing," IEEE Trans. Microwave Theory Tech. **43**, 2144–2150 (1995).

[10] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," Ann. Stat. **1**, 209–230 (1973).

[11] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," J. Am. Stat. Assoc. **90**, 577–588 (1995).

[12] S. N. MacEachern and P. Müller, "Estimating mixture of dirichlet process models," J. Comput. Graph. Stat. **7**, 223–238 (1998).

[13] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," J. Am. Stat. Assoc. **101**, 1566–1581.

[14] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," Technical Rep. No. 653, U.C. Berkeley Statistics, 2004.

[15] P. Runkle, L. Carin, L. Couchman, T. Yoder, and J. Bucaro, "Multi-aspect target identification with wave-based matched pursuits and continuous hidden Markov models," IEEE Trans. Pattern Anal. Mach. Intell. **21**, 1371–1378 (1999).

[16] J. Sethuraman, "A constructive definition of Dirichlet priors," Stat. Sin. **4**, 639–650 (1994).

[17] D. Blackwell and J. B. MacQueen, "Ferguson distributions via polya urn schemes," Ann. Stat. **1**, 353–355 (1973).

[18] C. E. Rasmussen, "The infinite Gaussian mixture model," Adv. Neural Inf. Process. Syst. **12**, (2000).

[19] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE **77**, 275–285 (1989).

[20] N. Dasgupta, P. Runkle, L. Carin, L. Couchman, T. Yoder, J. Buraco, and G. Dobeck, "Class-based target identification with multi-aspect scattering data," IEEE J. Ocean. Eng. **28**, 271–282 (2003).

[21] A. Stolcke and S. Omohundro, "Hidden Markov model induction by Bayesian model merging," Adv. Neural Inf. Process. Syst. **5** (1993).

[22] M. J. Beal, Z. Ghahramani, and C. Rasmussen, "The infinite hidden markov model," Adv. Neural Inf. Process. Syst. **14** (2002).

[23] J. Pitman, "Random discrete distributions invariant under size-biased permutation," Adv. Appl. Probab. **28**, 525–539 (1996).

[24] C. Antoniak, "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," Ann. Stat. **2**, 1152–1174 (1974).

[25] R. M. Gray, "Vector quantization," IEEE ASSP Mag. **1**, 4–29 (1984).

Noise propagation in linear and nonlinear inverse scattering

Dilip N. Ghosh Roy, John Roberts, and Matthias Schabel[a]
Utah Center for Advanced Imaging Research, University of Utah, 729 Arapeen Drive, Salt Lake City, Utah 84108

Stephen J. Norton
Fitzpatrick Institute for Photonics, Duke University, Durham, North Carolina 27708

The propagation of noise from the data to the reconstructed speed of sound image by inverse scattering within the framework of the Lippmann-Schwinger integral equation of scattering is discussed. The inversion algorithm that was used consisted in minimizing a Tikhonov functional in the unknown speed of sound. The gradient of the objective functional was computed by the method of the adjoint fields. An analytical expression for the inverse scattering covariance matrix of the image noise was derived. It was shown that the covariance matrix in the linear x-ray computed tomography is a special case of the inverse scattering matrix derived in this paper. The matrix was also analyzed in the limit of the linearized Born approximation, and the results were found to be in qualitative agreement with those recently reported in the literature for Born inversion using filtered backpropagation algorithm. Finally, the applicability of the analysis reported here to the obstacle problem and the physical optics approximation was discussed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713671]

I. INTRODUCTION

Inverse scattering[1–4] is an intensely researched area in current applied mathematics, and has wide applications in almost all branches of modern science and technology, from medical imaging and nondestructive evaluation of materials to problems in astronomy. The general objective is to recover an unknown physical object (scatterer and/or inhomogeneity) from a set of appropriately defined data. The data in inverse scattering consist of fields scattered by the unknown object under wave excitation. The framework of analysis in this work is the famous *Lippmann-Schwinger* integral equation of scattering.[1–4] Let T denote the integral operator of scattering. If the scatterer is "weak," then the uniform norm of T, $\|T\|_\infty \ll 1$, where $\|T\|_\infty = (1/2)(k_0 a)^2 \|1-n\|_\infty$. $1-n$ is the deviation of the scatterer refractive index n relative to that of the background which was assumed to be unity, $k_0 = 2\pi/\lambda$ is the wave number of the exciting wave which is of wavelength λ. a is the dimension of the scatterer. Moreover, $\|x(t)\|_\infty = \max|x(t)|$ over a certain interval of t and x in a suitable space X. In this case, the inverse scattering problem can be linearized by considering only the single scattering events. This is the well-known *Born* and *Rytov* approximation.[5–7] In these approximations, the incident wave is assumed to undergo a phase shift of less than π in traversing the scatterer. In general, however, the scatterer cannot be assumed to be weak. Also, in practice, scattering is frequently in the *resonance region* in which the wavelength is less than or comparable to the characteristic size of the scatterer.[1] Under such conditions, $\|T\|_\infty$ cannot be considered to be small, and inverse problems in the resonance region are improperly posed

and nonlinear. The multiple scattering interactions between the propagating wave and the inhomogeneity must be taken into account. Linearization is, therefore, not possible, and the inverse solutions are usually obtained by iteratively minimizing a suitably constructed, regularized objective function.

The scattered fields measured at the detectors are always corrupted by noise, the sources of which are physical, but can also arise from the numerics. The reconstructed image is, therefore, also noisy. Moreover, the propagation of noise depends upon the mathematical procedure by which the data are transformed into the image, that is, the reconstruction algorithm. Thus the understanding of how noise in the data propagates to the image via a given reconstruction algorithm is clearly an important problem, not only for understanding the characteristics of the noise in the reconstructed images produced by the algorithm, but also for comparing the performances of reconstruction algorithms for given noise characteristics in the measurements. The propagation of noise under the weak scattering condition, where the linearized Born or Rytov approximation holds, has been investigated by several researchers.[8–11] In this paper, the attention is focused on the nonlinear inverse scattering of an arbitrary scatterer with frequencies in the resonance region.

The paper is organized as follows. The direct problem and its Fréchet differentiability are briefly reviewed in Sec. II. The inversion algorithm and the calculation of the gradient of a nonlinear functional via the *adjoint field*[12–14] are presented in Sec. III. The inverse scattering covariance matrix of the image noise is derived in Sec. IV, and its relation to that of the linear x-ray computed tomography appears in Sec. V. In Sec. VI, the inverse scattering covariance matrix is specialized to the limiting case of the Born approximation. Section VI also includes a short discussion about the applicability of the present analysis to obstacle problems and

---

physical optics approximation. Finally, a brief summary of the work is presented in Sec. VII.

II. THE DIRECT PROBLEM AND FRÉCHET DIFFERENTIABILITY

The scattering is described by the *Lippmann-Schwinger* integral equation,[1–4] which is

$$\psi(x) = \psi^{\text{in}}(x) - k_0^2 \int_\Omega G^{(0)}(x|y)\gamma(y)\psi(y), \quad \Omega \in R^d,$$
$$d = 2,3, \tag{1}$$

subject to Sommerfeld's *radiation condition*:

$$\lim_{|x|\to\infty} |x|^m [\partial_{|x|}\psi^{\text{sc}} - ik_0\psi^{\text{sc}}] = 0,$$

$m = (d-1)/2$, and the limit is uniform in $|x|$ in all directions, where d is the dimensionality of the problem. Ω is the compact support of γ, that is, a bounded region in space. $\psi^{\text{in}}(x) = e^{ik_0\hat{k}_0 \cdot x}$, $\hat{k}_0 \in S^{d-1}$, the unit sphere in R^d, is the incident plane wave. $\psi = \psi^{\text{in}} + \psi^{\text{sc}}$ is the *total field*, ψ^{sc} being the scattered component. From Eq. (1).

$$\psi^{\text{sc}}(x) = -k_0^2 \int_\Omega G^{(0)}(x|y)\gamma(y)\psi(y), \quad x \notin \Omega. \tag{2}$$

In Eqs. (1) and (2), $\gamma(x) = 1 - (c_0/c(x))^2$ describes the spatial variation of the unknown inhomogeneity to be recovered, and the support of γ is assumed to be compact. $c(x)$ is the speed of sound in the object. $G^{(0)}(x|y)$ is the continuous free-space Green's function. $G^{(0)}(x|y) = (i/4)H_0^{(1)}(k_0|x-y|)$ in two dimensions (2D), and $G^{(0)}(x|y) = (1/4\pi)e^{ik_0|x-y|}/(|x-y|)$ in three dimensions. $H_0^{(1)}$ is the zeroth-order Hankel function.[15]

The solution of Eq. (1) can be written as

$$\psi = \mathcal{L}^{-1}\psi^{\text{in}}. \tag{3}$$

In Eq. (3), $\mathcal{L} = I + \mathcal{G}\gamma$: $L_2(\Omega) \to L_2(\Omega)$ is the Lippmann-Schwinger operator. I is the identity, and $\mathcal{G}$ the Green function operator. Furthermore, Eq. (2) can be recast as

$$F(\gamma) = \psi^{\text{sc}}. \tag{4}$$

The scattering operator $F(\gamma)$ in Eq. (4) is Fréchet differentiable. That is, there exists a linear operator, F', the Fréchet derivative of F, such that $\|F(\gamma+h) - F(\gamma) - F'h\| \sim o(\|h\|)$, $h \in L_2(\Omega)$ being an arbitrary vector. Some details about the existence and uniqueness of the Fréchet derivative can be found in Hohage.[16]

The mathematical form of the derivative is of importance to the discussions that follow. A terse derivation is, therefore, presented. Let $\tilde{\psi} = \gamma\psi$. Then $F(\gamma) = \mathcal{G}\tilde{\psi}$, and Eq. (1) becomes

$$\mathcal{L}^\dagger\tilde{\psi} = \gamma\psi^{\text{in}}, \tag{5}$$

in which $\mathcal{L}^\dagger = I - \gamma\mathcal{G}$ is the operator adjoint to $\mathcal{L}$. Also,

$$\psi = \psi^{\text{in}} + \mathcal{G}\tilde{\psi}, \tag{6}$$

Differentiating $F(\gamma) = \mathcal{G}\tilde{\psi}$ formally in γ yields: $F' = \mathcal{G}\tilde{\psi}'$. From Eqs. (5) and (6), it is not difficult to show that $\mathcal{L}^\dagger(\hat{\psi}'h) = \psi h$, from which it follows that $F' = \mathcal{G}[\mathcal{L}^\dagger]^{-1}\psi$.

In finite space dimensions, the operators $\mathcal{G}$, $\mathcal{L}$, and $\mathcal{L}^\dagger$ are replaced by their corresponding matrix operators, which are denoted here by G, L, and $L^\dagger$, respectively. We then obtain the Jacobian matrix of the scattered field, namely

$$J_{\text{sc}} = G[L^\dagger]^{-1}\Lambda_\psi, \tag{7}$$

and its transpose

$$J_{\text{sc}}^T = \Lambda_\psi L^{-1}G. \tag{8}$$

Λ is used to denote a diagonal matrix. Thus $\{\Lambda_\psi\}_{ij} = \psi_{lj}\delta_{ij}$. Let M be the number of the detectors, and the grid be of size $N \times N$. Then J_{sc} is a $M \times N^2$ matrix.

III. THE INVERSION ALGORITHM

The solution of the inverse scattering problem was obtained by minimizing the following nonquadratic, Tikhonov type functional:

$$\Phi(\gamma) = (1/2)(\psi^{\text{sc}} - \psi^{\text{me}})^T\Sigma^{-1}(\psi^{\text{sc}} - \psi^{\text{me}}) + \alpha\|\gamma - \gamma^*\|_2^2. \tag{9}$$

ψ^{me} is the measured (noisy) field, Σ the covariance matrix of the noisy measurements, and $\alpha > 0$, α small, the *regularization parameter*. γ^* is the initial estimate. It is known[17] that the regularizing term, as written in Eq. (9), contributes to the smoothness of the solution. It is to be noted that the functional, $\Phi(\gamma)$, is nonquadratic since the scattering operator, $F(\gamma)$ in Eq. (4) is nonlinear.

Let $\Phi_{,\gamma} = \nabla_\gamma\Phi$ be the functional derivative of Φ with respect to γ. The basic reconstruction is then given by the solution which is the null point of the gradient. However, in order to calculate the error propagation from data to image, it is necessary to specify the mathematical form of the functional derivative. In the present work, the derivative was obtained via the use of the adjoint fields. The result is[12]

$$\Phi_{,\gamma} = \text{Re}\sum_{j=1}^{N_W} [\Lambda_{\psi_j}\psi_j^\dagger] + 2\alpha(\gamma - \gamma^*), \tag{10}$$

where N_W is the total number of the waves incident on the scatterer, and $\psi_j^\dagger$ is the *adjoint field*. $\psi_j^\dagger$ is given by

$$\psi_j^\dagger = L^{-1}\psi_j^{\dagger\text{in}}, \quad \psi_j^{\dagger\text{in}} = G\overline{\tilde{r}}_j. \tag{11}$$

In Eq. (11), $\tilde{r}_j = \psi_j^{\text{sc}} - \psi_j^{\text{me}}$ is the complex residual vector at the detectors. Its mth component, $\psi_{jm}^{\text{sc}} - \psi_{jm}^{\text{me}}$, the complex residual at the mth detector, is

$$\tilde{r}_{jm} = [\text{Re}(r_{jm})/\text{Re}(\sigma_{jm}^2)] + i[\text{Im}(r_{jm})/\text{Im}(\sigma_{jm}^2)]. \tag{12}$$

Re(Im) is the real (imaginary) part of a complex variable, and the overbar implies complex conjugation.

A comparison of Eq. (11) with the Lippmann-Schwinger solution in Eq. (3) shows that $\psi_j^\dagger$ can be obtained from Eq. (1) if $\psi_j^{\dagger\text{in}}$ is substituted for ψ^{in} in that equation. Then from Eq. (10) it follows that the computation of the functional

derivative, $\Phi_{,\gamma}$, of the objective function, Φ in Eq. (9), requires the solutions of only two forward problems, one for ψ, and another for $\psi^\dagger$, for any incidence. It should be mentioned at this point that a closely related problem occurs in nonlinear tomography where the derivative of a nonlinear operator is calculated via the solution of a linear adjoint problem, called *adjoint differentiation*.[18]

The reconstruction algorithm, therefore, consists in solving the system: $\Phi_{,\gamma}=0$. The solution of the inverse problem then reduces to that of locating the zeros of a nonlinear system of equations given in Eq. (10). It is this algorithm that forms the basis for the analysis of the image noise in this work.

IV. THE PROPAGATION OF NOISE

Let $\xi=\langle\xi\rangle+\epsilon^\xi$ be a random quantity in which $\langle\xi\rangle$ is the ensemble mean and ϵ^ξ the noise component. Following this notation, $\gamma=\langle\gamma\rangle+\epsilon^\gamma$, $\psi=\langle\psi\rangle+\epsilon^\psi$, $\psi^{sc}=\langle\psi^{sc}\rangle+\epsilon^{sc}$, $\psi^\dagger=\langle\psi^\dagger\rangle+\epsilon^\dagger$, and $\psi^{me}=\langle\psi^{me}\rangle+n$. $\langle\psi^{me}\rangle$ is the noise-free data, and hence $\langle\psi^{me}\rangle=\psi^{sc}_{tr}$, ψ^{sc}_{tr} denoting the field scattered by the true object. Implicit is the assumption that the magnitudes of the fluctuations are small compared to their mean values. Equation (10) now becomes

$$\Phi_{,\gamma} = \mathrm{Re}[\langle\psi\rangle\langle\psi^\dagger\rangle + \langle\psi\rangle\epsilon^\dagger + \langle\psi^\dagger\rangle\epsilon^\psi]$$
$$+ [\text{higher order terms}] + 2\alpha(\langle\gamma\rangle - \gamma^*) + 2\alpha\epsilon^\gamma. \tag{13}$$

For the simplicity of notation, the sums over the index j for the incident field were omitted in Eq. (13). These will be introduced at the end. Setting the right-hand side of Eq. (13) to zero, and separating the mean and the random part, yields

$$\mathrm{Re}[\langle\psi\rangle\langle\psi^\dagger\rangle] + 2\alpha(\langle\gamma\rangle - \gamma^*) = 0,$$

for the ensemble mean, and

$$\mathrm{Re}[\langle\psi\rangle\epsilon^\dagger + \langle\psi^\dagger\rangle\epsilon^\psi] + 2\alpha\epsilon^\gamma = 0 \tag{14}$$

for the random part. We also introduce the assumption that the iterations have converged to within a δ neighborhood of the solution, in which case, the term $\langle\psi^\dagger\rangle\epsilon^\psi$ in Eq. (14) can be neglected, being of the higher order. Equation (14) then reduces to

$$\mathrm{Re}[\Lambda_{\langle\psi\rangle}\epsilon^\dagger] + 2\alpha\Lambda_\gamma = 0, \tag{15}$$

where $_{,\Lambda}$, as earlier, denotes a diagonal matrix.

Noting that $\langle\psi^\dagger\rangle \sim 0$ by virtue of the above-introduced assumption, the random part, $\epsilon^\dagger$, of the adjoint field, $\psi^\dagger$, is given by the relation: $\epsilon^\dagger=\langle L\rangle^{-1}\epsilon^{\dagger in}$, where $\epsilon^{\dagger in}$ is the random part of $\psi^{\dagger in}=G\bar{\bar{r}}$, as defined in Eq. (11). Moreover, $\langle L\rangle$ is the Lippmann-Schwinger operator in which $\gamma=\langle\gamma\rangle$, i.e., $\langle L\rangle=I-G\Lambda_{\langle\gamma\rangle}$. The quantity, $\bar{\bar{r}}$, defined in Eq. (12), contains both the data noise, $\bar{\bar{n}}$, as well as the noisy component of the scattered field. Let $\bar{\bar{r}}$ be written as: $\bar{\bar{r}}=\bar{\bar{r}}^{sc}+\bar{\bar{n}}$, where $\bar{\bar{r}}^{sc}$ denotes the latter. Then

$$\epsilon^\dagger = \langle L\rangle^{-1}G[\bar{\bar{r}}^{sc} + \bar{\bar{n}}], \tag{16}$$

where $\bar{\bar{r}}^{sc}$ is the noise component of the scattered part of $\bar{\bar{r}}$ and n is the data noise. From the definition in Eq. (12) for $\bar{\bar{r}}$:

$$\bar{\bar{r}}^{sc}_{jm} = \frac{\mathrm{Re}\,\bar{\epsilon}^{sc}_{jm}}{\mathrm{Re}\,\sigma^2_{jm}} - i\frac{\mathrm{Im}\,\bar{\epsilon}^{sc}_{jm}}{\mathrm{Im}\,\sigma^2_{jm}}, \tag{17}$$

and a similar expression applies to $\bar{\bar{n}}$. From Eqs. (8) and (16), one obtains

$$\mathrm{Re}[\Lambda_{\langle\psi\rangle}\epsilon^\dagger] = \mathrm{Re}[J^T_{sc}(\bar{\bar{r}}^{sc} + \bar{\bar{n}})]. \tag{18}$$

In order to proceed further, the scattered ϵ^{sc} is expressed in terms of the object ϵ^γ. It is readily verified that $\epsilon^{sc}=J_{sc}\epsilon^\gamma$. In view of Eq. (2),

$$\epsilon^{sc} = G\langle\gamma\rangle\epsilon^\psi. \tag{19}$$

Now replacing $\gamma=\langle\gamma\rangle+\epsilon^\gamma$ and $\psi=\langle\psi\rangle+\epsilon^\psi$ in the Lippmann-Schwinger Eq. (1), separating out the ensemble averaged and the random part on both sides of the result, it is readily obtained that

$$\epsilon^\psi = [\langle L\rangle^{-1}G\Lambda_{\langle\psi\rangle}]\epsilon^\gamma \tag{20}$$

to the first order. Substituting Eq. (20) in Eq. (19) then yields $\epsilon^{sc}=J_{sc}\epsilon^\gamma$, upon using Eq. (7). The identity also follows from the definition of the Jacobian matrix.

We next use the relation $\epsilon^{sc}=J_{sc}\epsilon^\gamma$, in Eq. (17), and replace the result in Eq. (18). Substituting the outcome in Eq. (15), and upon setting the result to zero yields

$$(\mathrm{Re}(J^T_{sc})\mathrm{Re}(\Sigma^{-2})\mathrm{Re}(J_{sc}) + \mathrm{Im}(J^T_{sc})\mathrm{Im}(\Sigma^{-2})\mathrm{Im}(J_{sc})$$
$$+ 2\alpha I)\epsilon^\gamma = \mathrm{Re}(J^T_{sc})\mathrm{Re}(\Sigma^{-2})\mathrm{Re}(\bar{n})$$
$$+ \mathrm{Im}(J^T_{sc})\mathrm{Im}(\Sigma^{-2})\mathrm{Im}(\bar{n}). \tag{21}$$

In Eq. (21), Σ denotes the covariance matrix of the data noise. Thus $\Sigma_{ij}-=\sigma_{ij}$.

In order to simplify the notations, let us define $P_i[J]=P_i\{[\Sigma^{-1}][J_{sc}]\}$, and $P_i[J^T]=P_i\{[J^T_{sc}]P_i[\Sigma^{-1}]\}$, $i=1,2$. The subscripts on P are used to indicate the real and imaginary parts of the operator. Thus $P_{i=1}(P_{i=2})$ imply that only the real (imaginary) parts of the operators in the argument of P are to be considered. Moreover, we now sum over the number of incident waves N_W in the above-presented expressions, and define $A=\Sigma_{j=1}^{N_W}\Sigma_{i=1}^2 P_i[J^T_j]P_i[J_j]$, and $B=\Sigma_{j=1}^{N_W}\Sigma_{i=1}^2 P_i[J^T]P_i[\bar{n}_j]$. Equation (21) then takes the form: $[A+2\alpha I]\epsilon^\gamma=B\bar{n}$, from which

$$\epsilon^\gamma = [A + 2\alpha I]^{-1}B\bar{n}. \tag{22}$$

Now $\mathrm{cov}(\gamma)=\langle\epsilon^\gamma\{\epsilon^\gamma\}^T\rangle$. From Eq. (22), it follows immediately that

$$\mathrm{cov}(\gamma) = [A + 2\alpha I]^{-1}\langle B\bar{n}\{B\bar{n}\}^T\rangle[[A + 2\alpha I]^T]^{-1}. \tag{23}$$

For a nonstochastic scatterer such as an acoustic inhomogeneity, it is reasonable to assume that data are uncorrelated with respect to the receiving transducers and view angles. The scattering data are, however, complex, and their real and imaginary parts are both corrupted by noise. This is

expressed through the complex σ. A correlation can exist between σ_{Re} and σ_{Im}. However, for a nonrandom acoustical scatterer, it is reasonable to neglect this correlation (see also Ref. 11). In other words, $cov(n_{jm}, n_{j'm'}) = \sigma^2 \delta_{jj'} \delta_{mm'}$. The covariance matrix in Eq. (23) then reduces to

$$cov(\gamma) = [A + 2\alpha I]^{-1} A[[A + 2\alpha I]^{-1}]^T. \qquad (24)$$

Equation (24) is the covariance matrix of the image noise in the inverse scattering algorithm given in Eq. (10).

V. LINEAR PROBLEM AND REAL DATA: X-RAY CT

The P operators in Sec. IV are nonlinear, i.e., they depend nonlinearly on the solution $\langle \gamma \rangle$. This is explicitly indicated in the fundamental equation of scattering, namely, $\psi^{sc} = F(\gamma)$, [Eq. (4), Sec. II]. $F(\gamma)$ indicates that F is a nonlinear functional of γ. Contrarily, if the operator F acts linearly on γ, as for example, if F is a matrix acting on a vector γ, then it is written as F_γ. In Eqs. (7) and (8) for the Jacobians, the nonlinearity is expressed through $\Lambda_{\langle \psi \rangle}$ and $\langle L \rangle$, both of which are $\langle \gamma \rangle$-dependent. The complex nature of the scattered data introduces a further layer of complication. Clearly, the covariance matrix must simplify if the operator is linear and the data g are real, that is, $g = F_\gamma$ instead of $g = F(\gamma)$ as in Eq. (4).

An important case of a linear problem with real data is that of the x-ray computed tomography.[19,20] In x-ray CT, the scattering matrix F becomes the projection matrix H, the ijth element of which represents the intersection of the jth x ray with the ith resolution element. The noisy x-ray CT data can be written as $g = H\gamma + n = \langle g \rangle + n$. In this case, Eq. (24) for $cov(\gamma)$ simplifies to

$$cov(\gamma) = [H^T \Sigma^{-2} H + 2\alpha]^{-1}[H^T \Sigma^{-2} H]$$
$$\times [[H^T \Sigma^{-2} H + 2\alpha]^{-1}]^T. \qquad (25)$$

A special case of Eq. (25) is the so-called MAP (*maximum a posteriori*) x-ray CT reconstruction algorithm[21] in maximum entropy with Poisson noise in the data. The corresponding objective function is given by

$$\Phi(\gamma) = \log\{p(g|\gamma)\} + \alpha \mathcal{R}(\gamma), \qquad (26)$$

where $\mathcal{R}(\gamma)$ is a positive definite symmetric matrix regularizer. $p(g|\gamma)$ is the likelihood of data g given a distribution γ, assumed to be positive. For Poisson noise,

$$p(g|\gamma) = \prod_{j=1}^{MN}\left[\exp\left(\sum_{i=1}^{N}H_{ji}\gamma_i\right)\right]\frac{(\sum_{i=1}^{N}H_{ij}\gamma_i)^{g_j}}{g j!}. \qquad (27)$$

Note that the quantity, g_j, in the denominator in Eq. (27) is an integer number, being the number of counts. MN is the total number of data, N being the size of the computational domain, and M the number of view angles. Carrying out the calculations of $\Phi_{,\gamma}$, with the *log-likelihood function*, $\log p(g|\gamma)$, $p(g|\gamma)$ in Eq. (27), (for details, see Ref. 22), it is obtained that

$$[H^T \Sigma^{-2} H + \alpha^2 R]\epsilon^\gamma = H^T \Sigma^{-2} n. \qquad (28)$$

From Eq. (28), the covariance for x-ray CT follows immediately. It is to be noted that for Poisson noise, Σ is the same as $\Lambda_{\langle g \rangle}$.

A comparison of Eqs. (24) and (28) shows that the error, ϵ^γ, has the same mathematical form in both x-ray CT and acoustic inverse scattering, namely,

$$\epsilon^\gamma = [C + R(\alpha)]^{-1}En,$$

where $R(\alpha)$ is the regularizing term. However, in the case of linear x-ray CT, C and E involve the linear operator of projection, H, and its transpose. In acoustic inverse scattering, however, C and E are relatively more involved, being nonlinearly dependent in γ. Further details and an application of Eq. (28) to a dynamic SPECT problem appear in Ref. 22, and some related results can be found in Refs. 23–25.

VI. THE LIMIT OF WEAK SCATTERING

In this section, Eq. (24) for the inverse scattering covariance matrix is analyzed in the limit of weak scattering in which the linearized Born approximation holds. Let us first assume that $Re\, \sigma_{jm} = Im\, \sigma_{jm} = \sigma$, $\forall j$, m, and therefore, $Re\, \Sigma^2 = Im\, \Sigma^2 = \sigma^2 I$. Σ^2 is a diagonal matrix: $(\Sigma^2)_{ii} = \sigma_i^2$. From now onward, attention will be confined to two dimensions.

Let

$$\mathcal{J} = \sum_{j=1}^{N_W} Re[J_j^T \bar{J}_j]. \qquad (29)$$

In the Born approximation, Eq. (8) reduces to

$$J_j^T = {}^{in}\Lambda_j G. \qquad (30)$$

${}^{in}\Lambda_j$ is the diagonal matrix for the jth incidence. That is, $({}^{in}\Lambda_j)_{kk} = \exp(ik_0 \hat{d}_j \cdot x_k)$, $\hat{d}_j$ being the unit vector specifying the direction of the jth incident wave. Also, the far-field Green's function has the form:

$$G_{k\ell} = \frac{e^{ik_0 R}}{\sqrt{R}}\left[\frac{ia}{2}e^{-i\pi/4}\sqrt{\pi k_0}J_1(k_0 a)\right]e^{-ik_0\hat{d}_\ell \cdot x_k}, \qquad (31)$$

in which use was made of the asymptotic expression:[1]

$$H_n^{(1)}(t) = \sqrt{\frac{2}{\pi t}}e^{i(t - n\pi/2 - \pi/4)}\left\{1 + \left(\frac{1}{t}\right)\right\}, \quad t \to \infty .$$

In Eq. (31), $\hat{d}_\ell$ is the unit vector in the direction of the ℓth detector and R is the radius of the detector ring. The computational domain was assumed to have been discretized into N^2 circular resolution elements following the widely used discretization scheme of Richmond.[26] The quantity, a, in Eq. (31), is the radius of such a circular pixel in the present, 2D, case. The geometry of the so discretized computational domain is shown in Fig. 1. In Fig. 1, x_k is the coordinate of the center of the kth circle in the grid, and $\hat{d}_j$, $\hat{d}_\ell$ are the unit vectors along the jth incident wave and the ℓth detector, respectively. The detectors are located on a ring of radius R surrounding the scatterer.

From Eqs. (30) and (31), we obtain the Jacobians in the Born approximation, namely
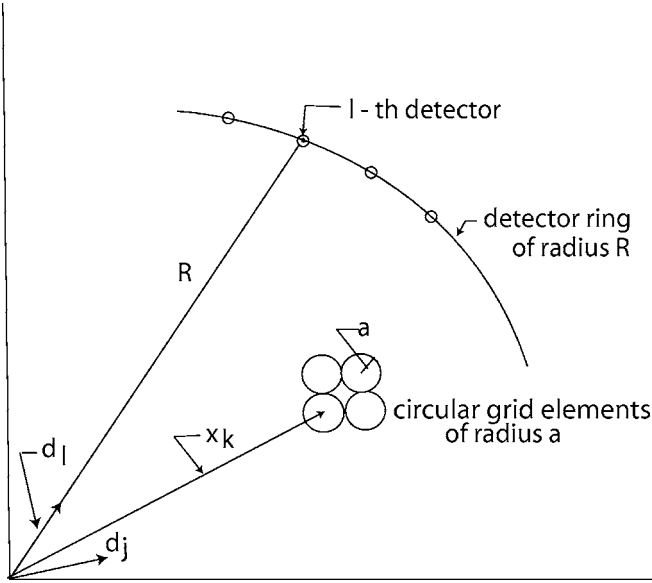
FIG. 1. A schematic of the geometry for the Born calculation illustrating the discretization of the computation grid into circular region. x_k is the location of the center of the kth resolution element in the grid. $\hat{d}_j$, $\hat{d}_\ell$ are the unit vectors along the jth incident wave and the ℓth detector, respectively. The detectors are located on a ring of radius R surrounding the scatterer. In the numerical computation in this work, the grid was discretized into 32×32 circular elements each of radius a. R/a was 200. That is, the detectors were placed on a radius which was 200 times the radius of a circular pixel.

$$[J_j^T]_{km} = \frac{e^{ik_0 R}}{\sqrt{R}}\left[\frac{ia}{2}e^{-i\pi/4}\sqrt{\pi k_0}J_1(k_0\alpha)\right]e^{ik_0(\hat{d}_j-\hat{d}_m)\cdot x_k}, \quad (32)$$

and

$$[\bar{J}_j]_{m\ell} = \frac{e^{-ik_0 R}}{\sqrt{R}}\left[-\frac{ia}{2}e^{i\pi/4}\sqrt{\pi k_0}J_1(k_0\alpha)\right]e^{-ik_0(\hat{d}_j-\hat{d}_m)\cdot x_\ell}. \quad (33)$$

Replacing Eqs. (32) and (33) in Eq. (29), and after some straightforward algebra, it follows that

$$\mathcal{J}_{k\ell} = C\left[\operatorname{Re}\sum_{j=1}^{N_W}\sum_{m=1}^{M}e^{-ik_0(\hat{d}_j-\hat{d}_m)\cdot(x_k-x_\ell)}\right], \quad (34)$$

where the constant $C=(\pi k_0 a^2/2R)[J_1(k_0 a)]^2$. Note that Eq. (34) reflects the well-known fact that in the Born approximation the Fourier frequencies of the object are confined within a circle (in 2D) of radius $2k_0$, and therefore, the maximum frequency that can be present in the recovered object is bounded above by $2k_0$.

Typically in practice, the product $N_W M$ is a large number for any realistic discretization of the computational domain. For example, for a 64×64 size of the computational grid, this product is 4096 at the minimum. In practical computations, this number is often larger as the system is almost always overdetermined in actual numerics. We, therefore, approximate the sums in Eq. (34) by integrals. Moreover, the sums over j and m are independent sums. Assuming without any loss in generality that $N_W=M$, each individual sum can be approximated by
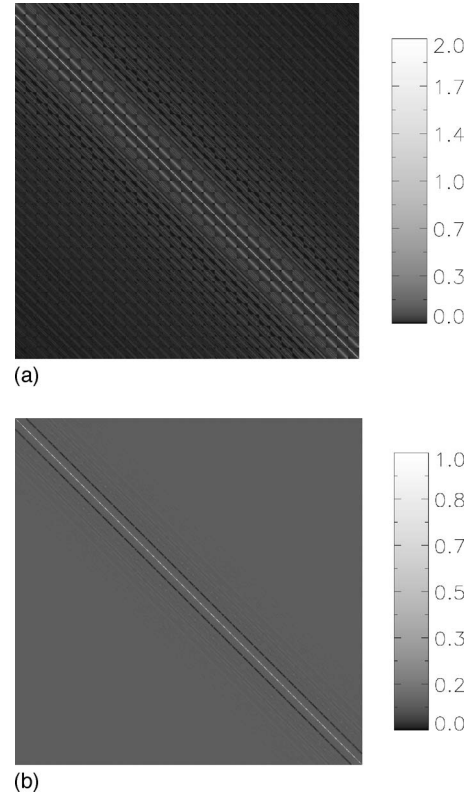
FIG. 2. (a) The 1024×1024 S matrix [Eq. (36)] for a 32×32 grid. (b) The corresponding T matrix with $\alpha=10^{-4}$. The block Toeplitz block structure of the matrices are visible.

$$\sum_{j=1}^{N_W}\cos(z\cos\phi) \sim \frac{N_W}{\pi}\int_0^{\pi}\cos(z\cos\phi)d\phi = N_W J_0(z), \quad (35)$$

recalling[15] that the integral in Eq. (35) is one of the representations of the cylindrical Bessel function J_0. Also, in Eq. (35), $z=k_0|x_k-x_\ell|$.

In view of Eq. (35), Eq. (24) for the covariance reduces to the following expression in the Born approximation, namely

$$[cov^B(\gamma)] = \sigma^2[T^{-1}ST^{-1}]. \quad (36)$$

In Eq. (36), $S=CN_W^2\mathcal{Z}$, $\mathcal{Z}_{k\ell}=J_0^2(k_0|x_k-x_\ell|)$, and the constant C was defined earlier. Moreover, $T=S+2\alpha I$. Note that S and T are each a block Toeplitz Toeplitz block or a BTTB matrix.[27]

It should be mentioned that the object enters the calculation through the noise in the data. In x-ray CT, for example, the object information is contained in Σ^{-2} in Eq. (28). Similarly, in the Born approximation, under the assumptions made in the derivation of Eq. (24) (essentially, the same assumptions also appear in Ref. 11), the object appears through σ^2 in Eq. (36) which may originate, say, from speckles. Thus the behavior of the covariance is determined completely by Eq. (36). It is illustrated here assuming a 32×32 computational grid.

The BTTB S matrix (1024×1024) is shown in Fig. 2(a), and the corresponding BTTB T matrix (1024×1024) is shown in Fig. 2(b) with the regularizing parameter $\alpha=10^{-4}$.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Roy *et al.*: Noise propagation in inverse scattering   2747
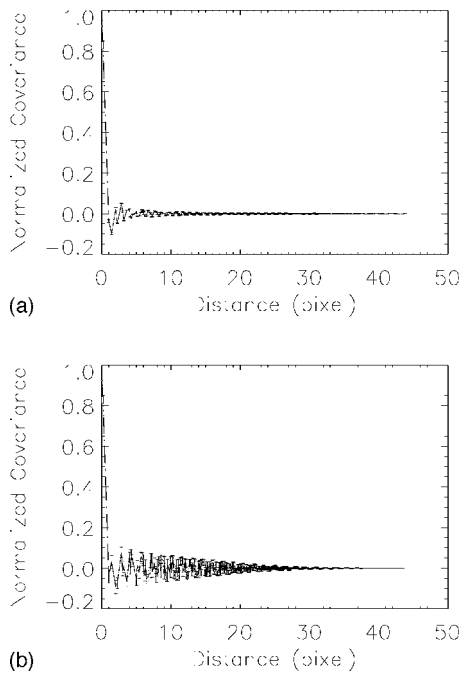
FIG. 3. The behavior of the covariance matrix, cov(γ) [Eq. (36)] with pixel separation. The variance or the diagonal elements of the matrix are constant. Panel (a) corresponds to the case in which $\lambda/a=2$, whereas (b) is for $\lambda/a=3$. Covariance is peaked at the center and falls off rapidly with the interpixel separation.

Toeplitz structures are clearly visible in Fig. 2. The covariance, cov$^B(\gamma)$, given in Eq. (36), is shown in Fig. 2. The variance is found to be constant in each pixel, whereas the off-diagonal elements of the covariance matrix are highly peaked around the center pixel, and decay rapidly in an oscillatory manner as the interpixel separation increases. In Fig. 3(a), λ/a was 2, while in Fig. 3(b), λ was three times the radius of a resolution element. In a recently published report,[11] the noise properties of Born objects were calculated using direct Born inversions. It is known[5–7] that Born data can be inferred by filtered backpropagation (FBPP), which is the analog of filtered backprojection of x-ray CT[19,20] in acoustical scattering. The behavior of the Born covariance matrix in Ref. 11 using FBPP inversion is qualitatively similar to that in Fig. 3. In both cases, the constant diagonal (the variance) and rapidly decaying oscillatory off-diagonal elements were obtained. Moreover, both methods produced σ^2 dependence in the covariance matrix. However, the dependence on k_0 is relatively more indirect in inverse scattering than in FBPP which produced a simple k_0^3 dependence. Since the inversion algorithms are different, only a qualitative agreement is to be expected.

Remark. The preceding analysis of the propagation of noise from the data to the reconstructed image was carried out in the framework of the Lippmann-Schwinger integral equation of scattering. The Lippmann-Schwinger equation provides the appropriate framework for scattering calculations from inhomogeneities, that is, when the scattering potential has no discontinuity across its support. However, the analysis can also be extended to the obstacle problem in which the potential is discontinuous across its boundary. For the obstacle problem, the appropriate framework of analysis

is the *Helmholtz representation*,[1,3] which involves surface integrals. However, the inversion can be formulated into a problem in nonlinear parameter estimation[1] involving objective functions which are essentially similar to Φ in Eq. (9). The parameters to be estimated in the case of the obstacles are the variables which parametrize the surface, e.g., spherical harmonics. The algorithm, Eq. (10), remains unchanged, and the functional derivative (now in the parametric variables of the surface) of the objective functional can again be obtained, as in the case of the inhomogeneities, via the method of adjoint fields, and again requiring the solutions of only two forward problems. The details of the procedure appear in Ref. 28. However, in the case of obstacle scattering, the derivatives are not the regular partial derivatives (as in Sec. III), but are the so-called *shape derivatives*,[29–33] which are defined in the following manner. If $V:R^3 \to R^3$ is a vector field that deforms a domain $\Omega_0 \in R^3 \to \Omega_t \in R^3$, t being the perturbation, then the first-order shape derivative of ψ, $\psi^{[1]}$, is defined by

$$\psi^{[1]} = \text{limit}_{t \to 0} \frac{\psi_t(I+tV)(X) - \psi_0(X)}{t}, \quad \forall\, X \in R^3.$$

ψ_0, ψ_t denote the fields for the unperturbed and perturbed boundary, respectively, corresponding to the boundary condition of the scattering problem, $\psi^{[1]}$ is the *total, material, or substantial derivative*. Moreover, $\psi^{[1]} = \psi' + V\cdot\nabla\phi$, ψ' being the partial derivative, namely, $\psi' = \text{limit}_{t \to 0}[\psi_t(X) - \psi_0(X)]/t$.

In a similar vein, the analysis can also be extended to the limit of the physical optics approximation. It is to be noted that the physical optics approximation bears the same relation to the Helmholtz representation as the Born approximation does to the Lippmann-Schwinger integral equation. In both cases, the total field within the surface integrals is replaced by the incident waves weighted by suitable constants.

VII. CONCLUSIONS

The propagation of noise from the data to the speed of sound image, reconstructed by inverse scattering within the framework of the Lippmann-Schwinger integral equation of scattering, was investigated. The inversion algorithm constituted in locating the zeros of the gradient of a Tikhonov-type functional in the unknown speed of sound. The gradient of the functional was computed by the method of the adjoint fields. An analytical expression was obtained for the image covariance matrix. When applied to the case of a linear mapping with real data, the inverse scattering expression reduced to that of the linear x-ray computed tomography. This was demonstrated by considering the maximum *a posteriori* likelihood algorithm using the maximum entropy functional. The full inverse scattering covariance was also investigated in the limiting case of the linearized Born approximation. It was found that in this linearized approximation, the variance remained uniform over the image, whereas the off-diagonal elements of the covariance matrix exhibited a fast decaying oscillatory behavior. Similar behaviors were also obtained in a recent report on Born inversion using the filtered backprojection algorithm. Finally, the applicability of the analysis presented to the problem of obstacle scattering and in the

limit of the physical optics approximation was briefly discussed.

ACKNOWLEDGMENTS

[1]D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer, Berlin, 1992).

[2]A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer, Berlin, 1986).

[3]D. N. G. Roy and L. S. Couchman, *Inverse Problems and Inverse Scattering of Plane Waves* (Academic, London, 2002).

[4]A. Ramm, *Multidimensional Inverse Problems* (Longman and Pitt, New York, 1992).

[5]A. J. Devaney, "Diffraction tomography," in *Inverse Methods in Electromagnetic Imaging, Part 2*, edited by W. M. Boerner, NATO ASI Series (Reidel, 1983), pp. 1107–1135.

[6]K. J. Langenberg, "Applied inverse problems for acoustic, electromagnetic and elastic wave scattering," in *Basic Methods of Tomography and Inverse Problems*, edited by P. C. Sabatier (Adam Hilger, Bristol, 1987).

[7]A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (SIAM, Philadelphia, 2001).

[8]D. Rouseff and R. Porter, "Diffraction tomography and stochastic inverse problem," J. Acoust. Soc. Am. **89**, 1599–1605 (1991).

[9]G. Tsihrintzis and A. Devaney, "Application of a maximum likelihood estimator in the experimental study in ultrasonic diffraction tomography," IEEE Trans. Med. Imaging **12**, 545–554 (1993).

[10]X. Pan, "A unified reconstruction theory for diffraction tomography with considerations of noise control," J. Opt. Soc. Am. A **15**, 2312–2326 (1998).

[11]M. Anastasio, M. A. Kupinski, and X. Pan, "Noise propagation in diffraction tomography: Comparison of conventional algorithms with a new reconstruction algorithm," IEEE Trans. Nucl. Sci. **45**, 2216–2223 (1998).

[12]S. J. Norton, "Iterative inverse scattering algorithms: Methods of computing Fréchet derivatives," J. Acoust. Soc. Am. **106**, 2653–2660 (1999).

[13]O. Dorn, H. Bertete-Aguirre, J. G. Berryman, and G. C. Papanocolau, "A nonlinear inversion method for 3d electromagnatic imaging using adjoint fields," Inverse Probl. **15**, 1523–1558 (1999).

[14]P. R. McGillivray and D. W. Oldenburg, "Methods for calculating Fréchet derivatives and sensitivities for the non-linear inverse problem: A comparative study," Geophys. Prospect. **38**, 499–524 (1994).

[15]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[16]T. Hohage, "On the numerical solution of a three-dimensional inverse medium scattering problem," Inverse Probl. **17**, 1743–1763 (2001).

[17]H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Ill-posed Problems* (Kluwer, Dordrecht, 1996).

[18]F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction* (SIAM, Philadelphia, 2001).

[19]H. H. Barrett and W. Swindell, *Radiological Imaging: The Theory of Image Formation, Detection, and Processing* (Academic, New York, 1996).

[20]F. Natterer, *The Mathematics of Computerized Tomography* (Teubner, Stuttgart, 1986).

[21]S. M. Kay, *Fundamentals of Statistical Signal Processing* (Prentice Hall, Englewood Cliffs, NJ, 1998).

[22]G. T. Gullberg, R. H. Huesman, D. N. G. Roy, J. Qi, and B. W. Reutter, "Estimation of the parameter covariance matrix for a one-compartment cardiac perfusion model estimated from a dynamic sequence reconstructed using map iterative reconstruction algorithms," in IEEE Nuclear Science Symposium and Medical Imaging Conference Record, Portland, Oregon, 2003, pp. 3019–3023.

[23]H. H. Barrett, D. W. Wilson, and B. M. W. Tsu, "Noise properties of the em algorithm I. Theory," Phys. Med. Biol. **39**, 833–846 (1994).

[24]J. Qi and R. M. Leahy, "Resolution and noise properties of map reconstruction for fully 3-D PET," IEEE Trans. Med. Imaging **19**, 493–506 (2000).

[25]J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," IEEE Trans. Image Process. **5**, 493–506 (1996).

[26]J. H. Richmond, "Scattering by a dielectric cylinder of arbitrary corss section shape," IEEE Trans. Antennas Propag. **AP-13**, 334–341 (1965).

[27]K. Chen, *Matrix Preconditioning Techniques and Applications* (Cambridge University Press, Cambridge, 2005).

[28]S. J. Norton, "Iterative algorithms for computing the shape of a hard scattering object: Computing the shape derivative," J. Acoust. Soc. Am. **116**, 1002–1008 (2004).

[29]O. Pironneau, *Optimal Shape Design for Elliptic Systems* (Springer, Berlin, 1984).

[30]A. Zochowski, *Mathematical Problems in Shape Optimization and Shape Memory Materials* (Lang, New York, 1992).

[31]O. Bruno and F. Reitich, "A new approach to the solution of problems of scattering by bounded obstacle," Proc. SPIE **122A**, 20–27 (1995).

[32]D. N. G. Roy, L. Couchman, and J. Warner, "Scattering and inverse scattering of sound-hard obstacles via shape deformation," Inverse Probl. **13**, 585–606 (1997).

[33]D. N. G. Roy, J. Warner, L. Couchman, and J. Shirron, "Inverse obstacle transmission problem in acoustics," Inverse Probl. **14**, 903–929 (1998).

Boundary element method for the acoustic characterization of a machine in bounded noisy environment

Christophe Langrenne, Manuel Melon,[a] and Alexandre Garcia
Laboratoire d'acoustique, Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75141 Paris cedex 3, France

In this article, a boundary element method is used to recover free field conditions from noisy bounded space situations. The proposed approach is based on the Helmholtz integral formulation. The method requires the knowledge of double layer pressure fields on two parallel closed surfaces surrounding the source. First, the outgoing and ingoing pressure field are separated using Helmholtz integral. Then, the incident field scattered by the tested source is subtracted from the outgoing field to recover the pressure field which would have been radiated in free space. To simplify the process, rigid body approximation for the source is used. The method is numerically tested in the following conditions: the chosen sound source is the upper spherical cap of a rigid sphere, the source is located at the center of a rigid spherical cavity, and a monopole secondary source is added to blur the primary pressure field. Simulations give good results for ka up to 5 when the discretization of the surfaces is sufficient. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2713670]

I. INTRODUCTION

Near field acoustic holography (NAH)[1] is a powerful tool for the identification and localization of vibratory sources. Different methods can be used to forward or backward propagate sound fields depending on the measurement geometry. For planar and cylindrical ones, Fourier acoustics, for example, is convenient and gives good results.[2] When dealing with small microphone arrays, statistical optimized near field holography applied by Steiner and Hald[3] in planar geometry can be used. This transfer matrix formulation based on the projection of the sound field on propagating and weighted evanescent waves does not use a spatial Fourier transform and thus reduces spatial aliasing. For arbitrary geometry, both boundary element method (BEM) and equivalent source methods allow the characterization of sound sources. BEM is particularly used at low frequencies to solve the inverse acoustical problem.[4] The equivalent source method, also called auxiliary source method or superposition method, is mainly used in far field for the calculation of directivity diagrams or radiated acoustical power because of the slowness of the series convergence in near field.[5,6] Hybrid methods of these two techniques have been developed to achieve results at higher frequencies without increasing processing time.[7,8]

As a no contact technique, NAH is well suited for industrial sound sources. However, it assumes that measurements are done in an anechoic room or sufficiently close to the machine so that no significant noisy field is added to the data. The inverse problem is known as "ill posed" and is very sensitive to noisy environment situations. Nevertheless, some machines cannot be stopped or displaced and measurements have to be done *in situ*. Then, measurement data also

contain direct sound generated by other machines and reverberated sound reflected by walls. Thus, localization of the vibrating parts of the studied machine cannot be achieved properly.

A solution consists of separating the convergent and divergent fields on a closed surface including the source. Weinreich and Arnold[9] have proposed a method based on the decomposition of double layer fields on spherical harmonics while Tsukernikov[10] uses an integral formulation to solve the problem. These formulations give good results while the wavelength is larger than the dimensions of the source. As an example, an integral formulation coupled to spherical harmonic decomposition method was used to evaluate field radiated by a subwoofer.[11] Reflections from walls of the semi-anechoic testing room below "cutoff frequency" were removed from data measurements and free field measurement conditions were recovered. Another example dealing with underwater acoustics is given by Romano *et al.*[12] which separates the incident field from the total pressure field to develop a virtual sonar.

However, diffraction of the convergent field on the measured object becomes significant at high frequencies. Then, one shall take into account the scattering effects on the sound source to improve the accuracy of the previous methods. A formulation based on a decomposition on spherical harmonics and on the transition matrix[13] of the machine has been developed by Bobrovnitskii[14] and has been successfully tested on simulated[15] or measured[16] data for simple configurations. For spherical targets, the transition matrix method quickly converges. Unfortunately, convergence speed of the harmonic series is rapidly deteriorated when increasing aspect ratio.[17] For complex source shapes, an experimental method has been developed by Bobrovniskii and Pavic[18] for measuring the transfer matrix coefficients. This approach al-

---

[a] author to whom correspondence should be addressed. Electronic mail: melon@cnam.fr
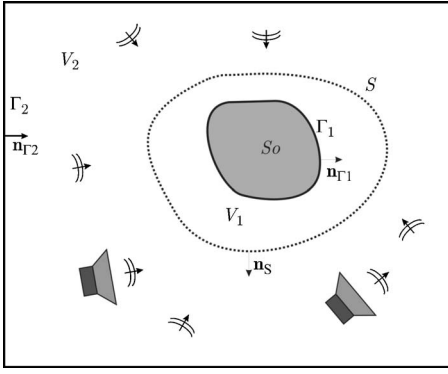
FIG. 1. Geometry of interest.

lows a quick estimation of the acoustical power radiated by a sound source but cannot give information on the near field pressure.

In this paper, we present a usable double layer BEM method for the separation of sound fields which take into account the diffraction on the sound source. Note that the two layers are closed surfaces surrounding the source. This method can been seen as a first process which has to be computed before resolving the acoustical inverse problem. Section II contains a description of the field separation and of the removal of the scattered field using rigid body assumption for the source. In Sec. III, a numerical simulation is presented. A vibrating spherical upper cap on an otherwise rigid sphere is used to insonify a rigid spherical cavity which contains a secondary source. We will summarize results and make suggestions for future research efforts in Sec. IV.

II. THEORY

The problem is to calculate the pressure field $p^f(\mathbf{r})$ radiated by a primary sound source in free space from double layer pressure measurements in a bounded noisy environment.

A. Problem definition

The geometry of interest is shown on Fig. 1. Let Γ_1 be the physical frontier of the sound source So to characterize. The domain V_1 in $\mathbb{R}^3$ is delimited by Γ_1 and the measurement surface S. The domain V_2, also in $\mathbb{R}^3$, is delimited by S and by the walls of the room Γ_2 and contains an arbitrary secondary sources distribution $Q(\mathbf{r})$. The pressure field $p(\mathbf{r})$ in $V = V_1 \cup V_2$ can be written as follows:

$$p(\mathbf{r}) = p^i(\mathbf{r}) + p^o(\mathbf{r}), \tag{1}$$

where $p^i(\mathbf{r})$ is the ingoing field radiated by all secondary sources and reflected by walls Γ_2 while $p^o(\mathbf{r})$ is the outgoing pressure field which is the superposition of the free field $p^f(\mathbf{r})$ and of the scattered field $p^s(\mathbf{r})$ on surface Γ_1

$$p^o(\mathbf{r}) = p^f(\mathbf{r}) + p^s(\mathbf{r}). \tag{2}$$

By using a $e^{-i\omega t}$ time dependence convention, the wave equation in V, for harmonic disturbances, can be written as follows:

$$\Delta p(\mathbf{r}) + k^2 p(\mathbf{r}) = -Q(\mathbf{r}), \tag{3}$$

where Δ is the Laplacian operator, k is the wave number, $Q(\mathbf{r})$ is an unknown distribution of sources (note that $Q(\mathbf{r}) = 0$ in V_1). Boundary conditions are given by

$$\begin{cases} \partial_{\mathbf{n}_S} p(\mathbf{r}) = ik\rho_0 c\upsilon(\mathbf{r}), & \mathbf{r} \in S \\ \partial_{\mathbf{n}_{\Gamma_i}} p(\mathbf{r}) = ik(\beta_i p(\mathbf{r}) + \rho_0 c\upsilon_\Gamma(\mathbf{r})), & \mathbf{r} \in \Gamma_i \\ & i = 1 \text{ or } 2, \end{cases} \tag{4}$$

where $\mathbf{n}_S$ and $\mathbf{n}_{\Gamma_i}$ are the respective normals to S, Γ_1 and Γ_2 surfaces, β_i is the specific admittance of surfaces Γ_1 or Γ_2, $\rho_0 c$ is the specific impedance of air, $\upsilon(\mathbf{r})$ is the normal particular velocity, and $\upsilon_\Gamma(r)$ is the normal driving velocity.

The standard Helmholtz integral formulation of the problem is written as follows for $\mathbf{r} \in V_i$ with $i = 1$ or 2:

$$p(\mathbf{r}) = \int_{V_i} G(\mathbf{r},\mathbf{r}')Q(\mathbf{r}')dV + \int_S \mathbf{n}_S \cdot [p(\mathbf{s}')\nabla G(\mathbf{r},\mathbf{s}')$$
$$- G(\mathbf{r},\mathbf{s}')\nabla p(\mathbf{s}')]dS + \int_{\Gamma_i} \mathbf{n}_{\Gamma_i} \cdot [p(\mathbf{s}')\nabla G(\mathbf{r},\mathbf{s}')$$
$$- G(\mathbf{r},\mathbf{s}')\nabla p(\mathbf{s}')]dS, \tag{5}$$

where $\mathbf{s}$ is a point on a boundary and $G(\mathbf{r},\mathbf{r}')$ is the free space Green's function for Helmholtz' equation

$$G(\mathbf{r},\mathbf{r}') = \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}. \tag{6}$$

B. Field separation

The field separation allows to calculate $p^i(\mathbf{r})$ and $p^o(\mathbf{r})$ from pressure $p(\mathbf{s})$ and pressure gradient $\partial_n p(\mathbf{s})$ on S. In practice, these quantities are calculated from pressure field data measured on two close surfaces: using mean values for $p(\mathbf{s})$ and approximation of Euler's equation by finite difference for $\partial_n p(\mathbf{s})$.[19]

The equations for the interior problem, i.e., V_1 domain, are

$$p^o(\mathbf{r}) - \int_S [p(\mathbf{s}')\partial_n G(\mathbf{r},\mathbf{s}') - G(\mathbf{r},\mathbf{s}')\partial_n p(\mathbf{s}')]dS$$
$$= \begin{cases} p(\mathbf{r}) \text{ for } \mathbf{r} \in V_1 \\ \left(1 - \frac{\Omega_S}{4\pi}\right)p(\mathbf{r}) \text{ for } \mathbf{r} \in S, \\ 0 \text{ for } \mathbf{r} \in V_2 \end{cases} \tag{7}$$

where $\Omega_S(\mathbf{r})$ is the solid angle coefficient given by[20]

$$\Omega_S(\mathbf{s}) = 4\pi + \int_S \partial_n\left(\frac{1}{|\mathbf{s}-\mathbf{s}'|}\right)dS. \tag{8}$$

Note that the contribution of the source So is represented by a surface integral on Γ_1 which corresponds to the outgoing pressure field $p^o(\mathbf{r})$

$$p^o(\mathbf{r}) = \int_{\Gamma_1} [p(\mathbf{s}')\partial_n G(\mathbf{r},\mathbf{s}') - G(\mathbf{r},\mathbf{s}')\partial_n p(\mathbf{s}')]dS. \tag{9}$$

As the pressure and pressure gradient are known on S, $p^o(\mathbf{s})$ can be calculated from Eq. (7)

$$p^o(\mathbf{s}) = \left(1 - \frac{\Omega_S}{4\pi}\right)p(\mathbf{s}) + \int_S [p(\mathbf{s}')\partial_n G(\mathbf{s},\mathbf{s}')$$
$$- G(\mathbf{s},\mathbf{s}')\partial_n p(\mathbf{s}')]dS. \qquad (10)$$

At low frequencies, i.e., when the wavelength is much larger than the dimensions of the sound source, scattering can be neglected, thus, $p^s(\mathbf{r}) \approx 0$. Then, the free field radiated by the machine is equal to the outgoing field and can be calculated from Eq. (10) and from double layer pressure measurements on S. In practice, this method can be used at low frequency in ordinary rooms or in an anechoic room below the cutoff frequency.

At high frequencies, scattering on the sound source is no longer negligible and must be taken into account. In this case, one needs to know the incident field on Γ_1 to calculate the diffracted field. The equations for the exterior problem, i.e., V_2 domain, are

$$p^i(\mathbf{r}) + \int_S [p(\mathbf{s}')\partial_n G(\mathbf{r},\mathbf{s}') - G(\mathbf{r},\mathbf{s}')\partial_n p(\mathbf{s}')]dS$$
$$= \begin{cases} 0 \text{ for } \mathbf{r} \in V_1 \\ \dfrac{\Omega_S}{4\pi}p(\mathbf{r}) \text{ for } \mathbf{r} \in S, , \\ p(\mathbf{r}) \text{ for } \mathbf{r} \in V_2 \end{cases} \qquad (11)$$

with the contribution of secondary sources and reflections on walls Γ_2 given by

$$p^i(\mathbf{r}) = \int_{V_2} G(\mathbf{r},\mathbf{r}')Q(\mathbf{r}')dV' + \int_{\Gamma_2} [p(\mathbf{s}')\partial_n G(\mathbf{r},\mathbf{s}')$$
$$- G(\mathbf{r},\mathbf{s}')\partial_n p(\mathbf{s}')]dS. \qquad (12)$$

Then using the extinction theorem of Eq. (11) yields to the following formula for $p^i(\mathbf{r})$ for $\mathbf{r} \in V_1$:

$$p^i(\mathbf{r}) = - \int_S [p(\mathbf{s}')\partial_n G(\mathbf{r},\mathbf{s}') - G(\mathbf{r},\mathbf{s}')\partial_n p(\mathbf{s}')]dS. \qquad (13)$$

The calculation of the scattered field is done in the next section using an integral formulation of the diffraction theory.

C. Removal of the scattered field

The calculation of the diffraction of the incident field on an elastic body is possible but requires the resolution of the fluid-structure interaction problem. However, this task is tedious to achieve by measurements means and is time consuming for finite element method modelization. Moreover, in many cases, contact or very near field measurements can be impossible to perform and modelization can be inaccurate for complex machines. In order to simplify the resolution of the problem, an uncoupled fluid-structure assumption can be done. Fortunately, most industrial sources have much lower admittances than air, i.e., $\beta_i \approx 0$ is admitted. Then, for source

So stopped ($p^f(\mathbf{r})=0$) and assuming rigid body for the structure source, Eq. (4) gives $\partial_{\mathbf{n}_{\Gamma_i}} p(\mathbf{r})=0$. The exterior problem to Γ_1 can be rewritten as follows:

$$p^i(\mathbf{r}) + \int_{\Gamma_1} p^b(\mathbf{s}')\partial_n G(\mathbf{s},\mathbf{s}')dS = \begin{cases} p^b(\mathbf{r}) \text{ for } \mathbf{r} \in V \\ \dfrac{\Omega_{\Gamma_1}}{4\pi}p^b(\mathbf{s}) \text{ for } \mathbf{s} \in \Gamma_1 \end{cases}, \qquad (14)$$

where $p^b(\mathbf{s})$ is defined as the blocked pressure. This approach turns the scattering problem into a radiation problem.[2] The integral term corresponds to the scattered pressure field in the rigid case. Finally, the scattered field on S is given by

$$p^s(\mathbf{r}) = \int_{\Gamma_1} p^b(\mathbf{s}')\partial_n G(\mathbf{r},\mathbf{s}')dS, \quad \mathbf{r} \in S. \qquad (15)$$

Using pressure field $p^i(\mathbf{r})$ computed from Eq. (13) for $\mathbf{r} \in \Gamma_1$, $p^b(\mathbf{s})$ is calculated from Eq. (14) in the rigid case. Finally, Eqs. (2), (10), and (15) yield to

$$\tilde{p}^f(\mathbf{r}) = p^o(\mathbf{r}) - p^s(\mathbf{r}), \quad \mathbf{r} \in S, \qquad (16)$$

where $\tilde{p}^f(\mathbf{r})$ is the free pressure field radiated by So with rigid body assumption.

III. SIMULATION

A. Discretization

Numerical solution of the integral equations developed in the previous chapter is based on its discretization on surface elements. An isoparametric formulation based on Seybert *et al.*[20] using quadratic shape functions (six node curvilinear triangular elements) has been implemented. Then, outgoing and ingoing fields are calculated using matrix formulations of Eqs. (10) and (13)

$$\mathbf{p}_S^o = \left[\mathbf{D}_S^S + \left(1 - \frac{\Omega_S}{4\pi}\right)\mathbf{I}\right]\mathbf{p}_S - [\mathbf{M}_S^S]\partial_n \mathbf{p}_S, \qquad (17)$$

$$\mathbf{p}_{\Gamma_1}^i = - [\mathbf{D}_S^{\Gamma_1}]\mathbf{p}_S + [\mathbf{M}_S^{\Gamma_1}]\partial_n \mathbf{p}_S, \qquad (18)$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{D}$ and $\mathbf{M}$ are, respectively, the dipole and monopole transfer matrices. The subscript letter stands for the integration surface while the superscript one corresponds to the target surface. The blocked pressure field on Γ_1 is calculated from Eq. (14) once Eq. (18) is solved

$$\mathbf{p}_{\Gamma_1}^b = \left[\frac{\Omega_{\Gamma_1}}{4\pi}\mathbf{I} - \mathbf{D}_{\Gamma_1}^{\Gamma_1}\right]^{-1} \mathbf{p}_{\Gamma_1}^i \qquad (19)$$

When the wave number k coincides with a Dirichlet eigenvalue, the solution is not unique. Then, the combined Helmholtz integral equation formulation (CHIEF) method, based on the extinction theorem, is used to overcome this problem.[21,22]

The scattered pressure on S is then obtained using

$$\mathbf{p}_S^s = [\mathbf{D}_{\Gamma_1}^S]\mathbf{p}_{\Gamma_1}^b. \qquad (20)$$

Langrenne *et al.*: Acoustic characterization in bounded noisy environment
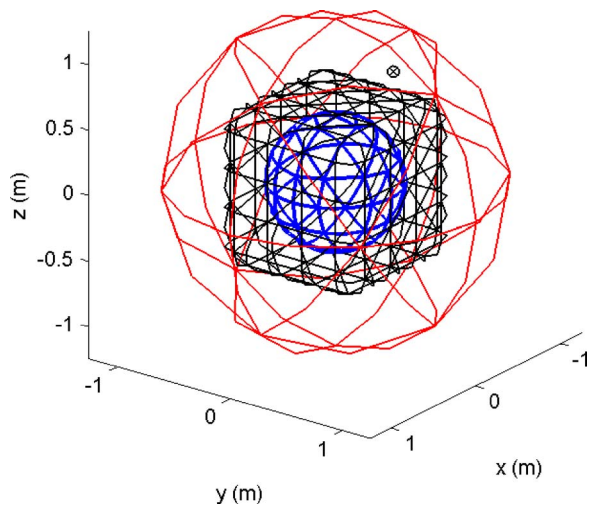
FIG. 2. (Color online) Geometry of the numerical experiment. Γ_1: smallest sphere, S: cube with smooth truncated edges, γ: largest sphere. The crossed circle shows the secondary source position.

B. Numerical experiment

The sound source is the vibrating upper spherical cap on an otherwise rigid sphere. The sphere, with a center O and a radius a of 0.5 m, is discretized with 258 points decomposed onto 128 curvilinear triangular elements. The upper spherical cap is defined by an angle θ_c of $\pi/9$ radians and vibrates with a velocity amplitude V_{sc} of 1 m/s. This mesh allows a maximum studying frequency of 388 Hz which corresponds to a 3.6 dimensionless wave number ka calculated with the well-known $\lambda/6$ criterion.

Three different cases are studied:

- First, the sound source is placed in free field conditions and the field separation process is applied on S. The measurement surface S is a cube with smooth truncated edges of 1.1 m and with several possible discretization ratios. This cube has the same center O than the sound source. This configuration allows checking the algorithms and evaluating the numerical errors: the calculated ingoing field must be negligible compared to the outgoing field.
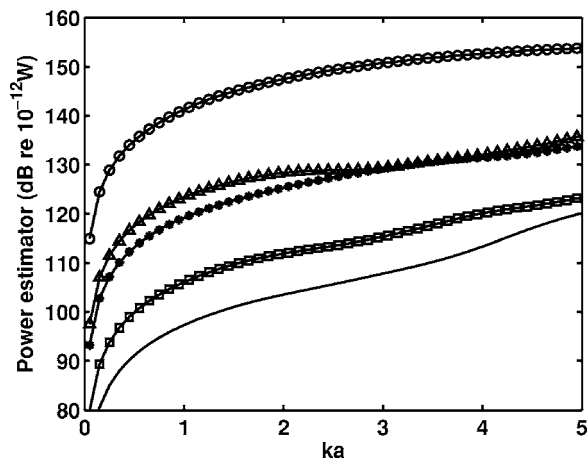
FIG. 3. Power estimator Π computed on S after separation process in free space. Outgoing field, $\bigcirc$: 594-node mesh. Ingoing field, $\triangle$: 90-node mesh, *: 210-node mesh, $\square$: 378-node mesh, solid line: 594-node mesh.
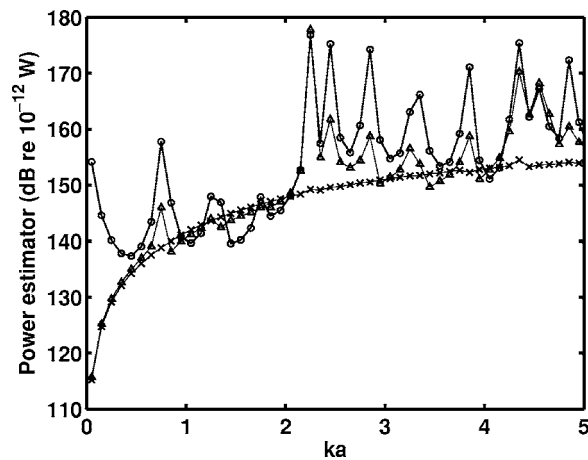
FIG. 4. Power estimator Π computed on S. $-\circ-$: total power. $-\triangle-$: outgoing power. $-\times-$ recovered free field power.

- For the second case, the sound source is placed in a concentric rigid sphere with a radius b of 1.25 m. The surface S is now meshed with 378 points decomposed onto 188 curvilinear triangular elements. Both field separation and removal of the scattering effects are computed.
- For the last case, a monopole secondary source is added at spherical coordinates $(1, \pi/7, 3\pi/5)$. For each frequency, monopole and spherical cap acoustical powers have been equalized. The geometry of this problem is sketched in Fig. 2.

Theoretical pressure fields are calculated using a 50th order spherical harmonic decomposition, which is more than sufficient at such frequencies. Equations allowing the calculation of the theoretical fields are given in the Appendix . For each case, the distance between the two simulated layers located on each side of S is 3 cm.

C. Results

The first configuration is computed to test the separation process. Calculations are performed without secondary source and in free space. In this case, the ingoing field should be null, but due to numerical processing errors, a nonphysi-
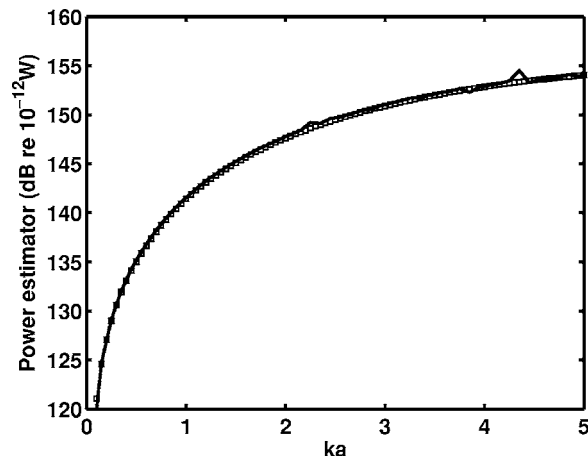
FIG. 5. Power estimator Π computed on S. $-\square-$: free field power. Thick line: recovered free field power.
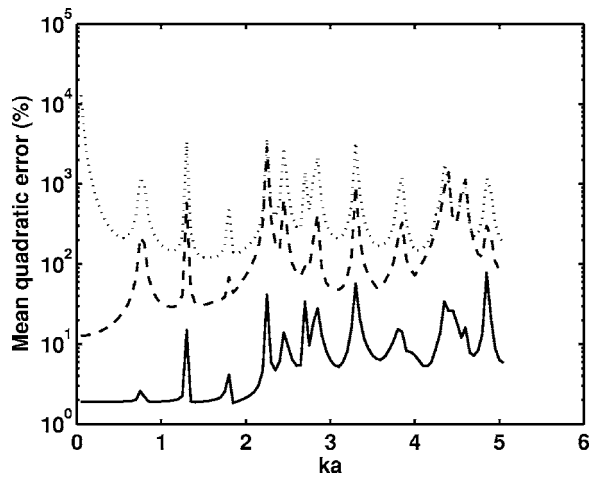
FIG. 6. Mean quadratic errors relative to the theoretical pressure field. Dotted line: total pressure field. Dashed line: outgoing pressure field. Solid line: recovered free pressure field.

$$\Pi = \int_S \frac{|p^e(s)|^2}{\rho_0 c} \, dS, \tag{21}$$

where p^e is the pressure field on which Π is estimated. It can either be the total pressure field $p(\mathbf{r})$, the outgoing field $p^o(\mathbf{r})$, or the recovered free pressure field $\tilde{p}^f(\mathbf{r})$.

In bounded space and with additional sound sources, Π is biased because all contributions are taken into account. In our case, this estimator allows to highlight the different stages of the proposed method.

When the number of nodes is larger than 378, the outgoing field power estimator is more than 30 dB higher that the ingoing field power estimator for ka up to 5. On the other hand, with small discretization ratios, the ingoing power could be only 15 dB smaller that the outgoing power. From this point, the 378-node configuration will be used for S.

For the second configuration, the source is placed in the spherical rigid cavity without secondary sources. In Fig. 4, different contributions of Π are plotted against ka. The total power estimator on S shows very strong fluctuations due to the modal behavior of the cavity. One can see that the rigid sphere strongly modifies the shape of the pressure field radi-

cal ingoing field is obtained. A power estimator Π is plotted in Fig. 3 for several discretizations of S (90, 210, 378 or 594 nodes). The estimator Π is computed using the quadratic pressure
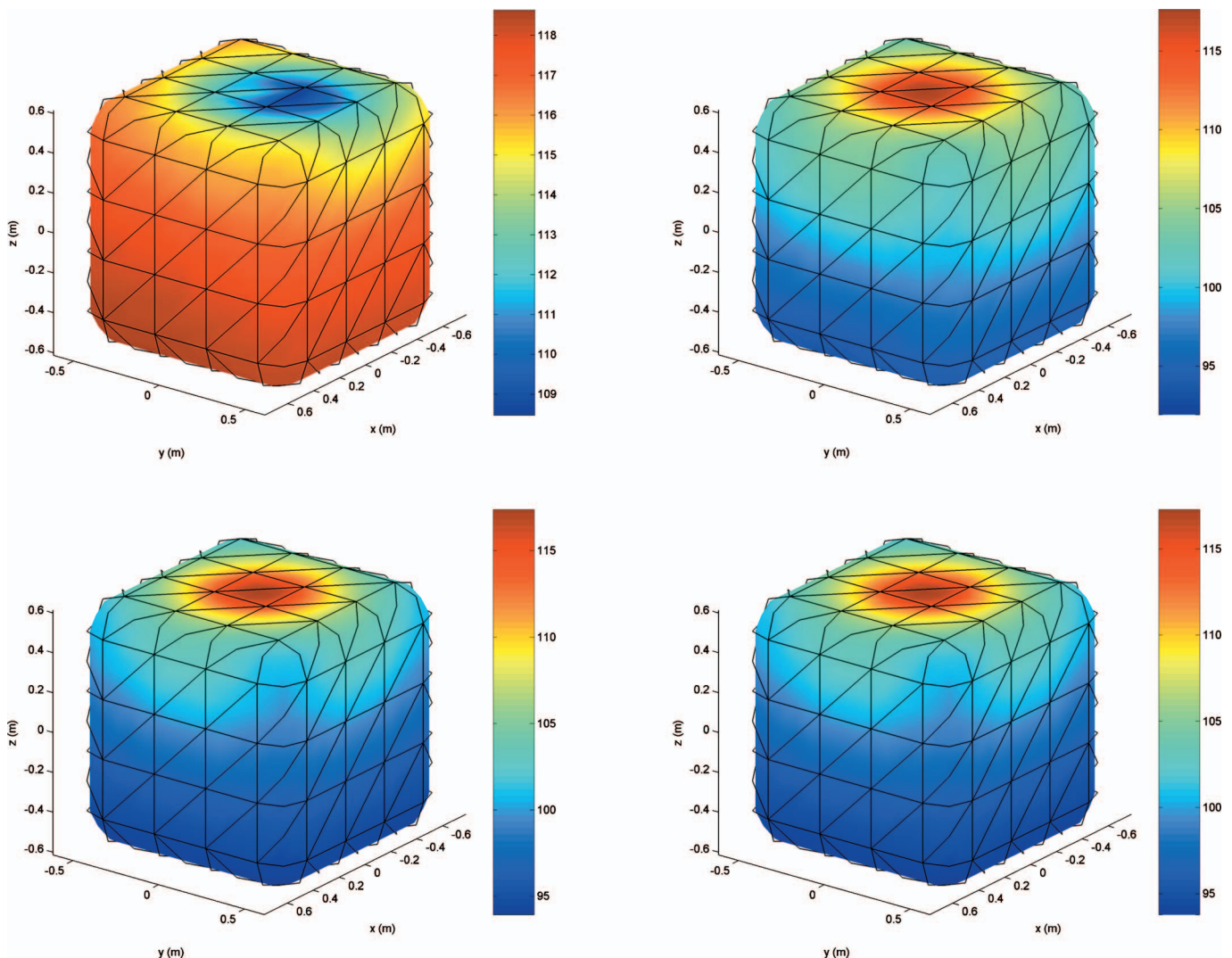
FIG. 7. Pressure field in dB on S for $ka=0.25$. (a) Total pressure field, (b) outgoing pressure field, (c) recovered free pressure field, (d) free pressure field.
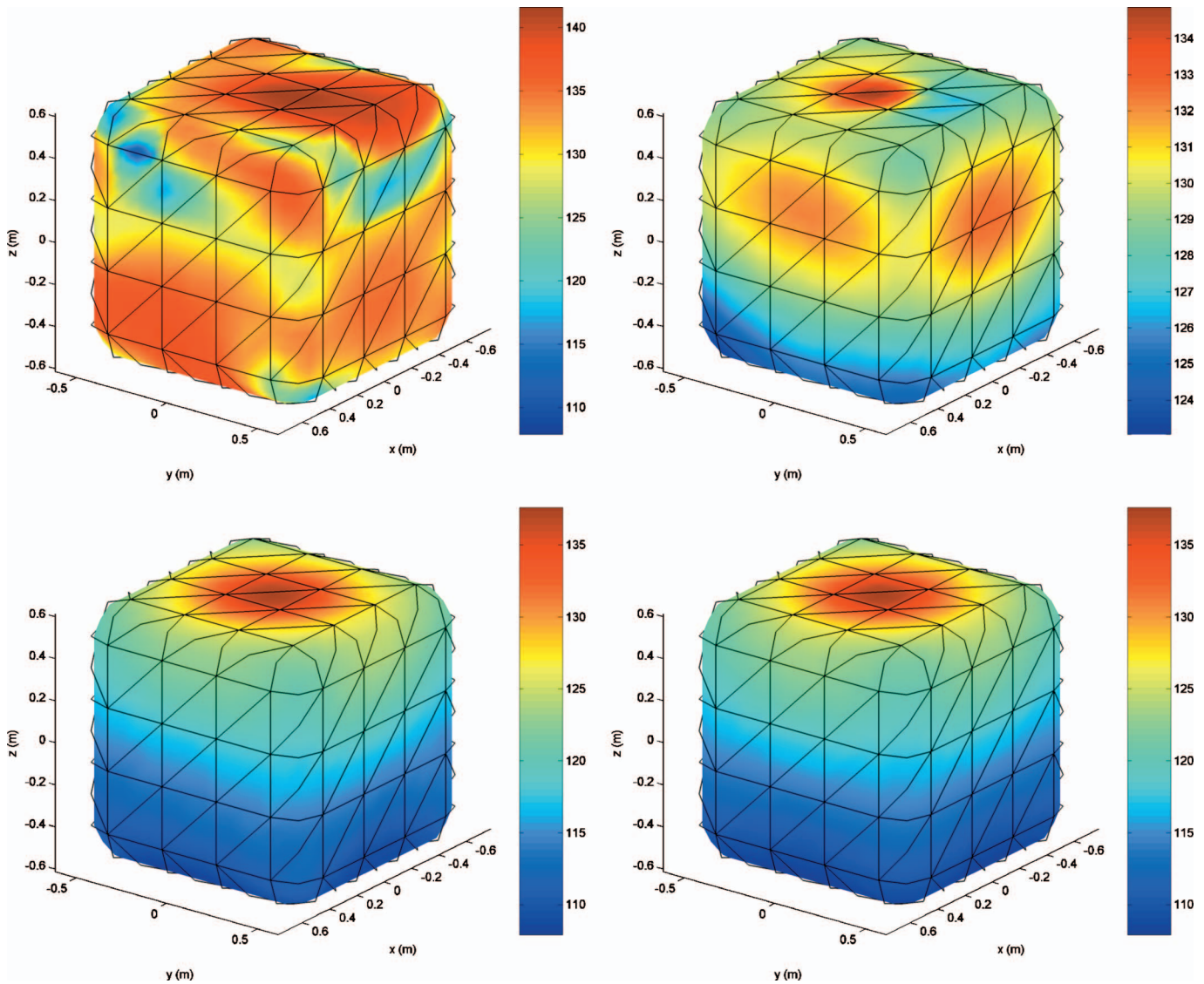
FIG. 8. Pressure field in dB on S for $ka=2.35$. (a) Total pressure field, (b) outgoing pressure field, (c) recovered free pressure field, (d) free pressure field.

ated by the spherical cap. Once the separation process is performed, the outgoing power curve is smoother but the shape is still very different from the one plotted in Fig. 3, particularly for the first cavity mode and for $ka>2$. Finally, the scattered field is subtracted from outgoing field to recover free space conditions. For $ka<0.5$, there are no differences between outgoing field and recovered free field conditions. For larger ka, resonances of the rigid cavity are accurately removed. A comparison, of the free field power and recovered free field power estimators is shown in Fig. 5. The agreement between these two curves is very good even for large ka. Nevertheless, one can see very small discrepancies (<1 dB) for some modal frequencies.

For the last simulation, the monopole source previously described is added in the cavity. The mean quadratic errors relative to the theoretical pressure in free field conditions are plotted versus ka in Fig. 6 for several pressure fields. The calculation is performed, on the discretized measurement surface, using the following equation:

$$E = \sqrt{\frac{\int_S |p^e(\mathbf{s}) - p^f(\mathbf{s})|^2 dS}{\int_S |p^f(\mathbf{s})|^2 dS}}. \tag{22}$$

As expected, the error between $p(\mathbf{r})$ and $p^f(\mathbf{r})$ is always high ($>100\%$). On the other hand, a simple separation of the outgoing field from the ingoing one gives error rates between $p^o(\mathbf{r})$ and $p^f(\mathbf{r})$ in the range of 11% for ka up to 0.5. This could be satisfactory for most industrial sources. For $ka<2.2$, only $\tilde{p}^f(\mathbf{r})$ can give error rates lower than 3% except for modal frequencies of the largest sphere. With a 378 node mesh and a $\lambda/6$ criterion, the maximum ka is equal to 2.7 on S. One can clearly notice a step in the error rate about this value. Thus, for $3<ka<5$, the minimal values of E are in the range of 5–7% and can go up to 60% at modal frequencies of the largest sphere.

Several maps of the pressure fields have been drawn to highlight the conclusions obtained on the error rate E. Figure

7 respectively shows total, outgoing, recovered free field and free field pressure levels for $ka=0.25$ on three sides of S. From the total pressure field [Fig. 7(a)], it is not possible to infer the position of the spherical cap which is on the upper part of Γ_1. After computation of the outgoing field [Fig. 7(b)], the spherical cap is then clearly located. At such low frequency, removal of the scattered field on Γ_1 is not mandatory. One can see that both calculated outgoing field [Fig. 7(b)] and recovered free field [Fig. 7(c)] are similar to the theoretical free field [Fig. 7(d)]. At higher frequencies, separation of the outgoing field from the ingoing field is not sufficient anymore. An illustration is given in Fig. 8 for $ka=2.35$. In this case, both separation process and removal of the scattering are required: only fields $p^f(\mathbf{r})$ and $\tilde{p}^f(\mathbf{r})$ in Fig. 8 appear nearly identical.

IV. SUMMARY

In this paper, a double layer BEM method for the measurement of sound sources in bounded noisy environment has been presented. First, theoretical background has been reviewed. The two steps of the processing, i.e., separation of the outgoing field from the ingoing field and removal of the scattered field on the sound source, have been detailed. Then, a numerical experiment has been performed to test out the method. It has been shown that for $ka<0.5$, separation alone gives good results. However, when dealing with higher frequencies, typically $ka>0.5$, diffraction on the sound source is no longer negligible. Then, both steps of the method should be processed to give an accurate estimation of the pressure field radiated by the sound source. The satisfactory results obtained here on simulations gives hope for the characterization of real sources in noisy reverberant spaces. Measurements are in progress in our laboratory and will be reported in the very near future. Note that, using a well-adapted regularization of the so-called inverse problem, this method could be coupled to a backward propagation process of the sound field on surface S to surface Γ_1. This additional calculation layer would be a great help when trying to locate the most vibrating parts of a machine.

APPENDIX

The pressure field $p_{sc}(\mathbf{r})$, radiated by the axisymmetric vibrating spherical upper cap on an otherwise rigid sphere located at the center of a rigid spherical cavity is given by the following equation:

$$p_{sc}(\mathbf{r}) = \sum_{n=0}^{\infty} [a_{n0}h_n^{(1)}(kr)Y_{n0}(\theta,\phi) + b_{n0}j_n(kr)Y_{n0}(\theta,\phi)],$$
(A1)

where Y_{n0} are the orthonormal spherical harmonics, $h_n^{(1)}$ is the first kind of spherical Hankel function and j_n is the spherical Bessel function. Note that the Hankel function corresponds to the outgoing field while the Bessel function corresponds to the regular field.

The amplitude coefficients a_{n0} and b_{n0}, respectively, are

$$\begin{cases} a_{n0} = i\rho_0 c V_n \dfrac{j_n'(kb)}{h_n'^{(1)}(ka)j_n'(kb) - h_n'^{(1)}(kb)j_n'(ka)} \\ b_{n0} = -i\rho_0 c V_n \dfrac{h_n'^{(1)}(kb)}{h_n'^{(1)}(ka)j_n'(kb) - h_n'^{(1)}(kb)j_n'(ka)} \end{cases},$$
(A2)

where V_n are the coefficients of the normal velocity expanded in terms of Legendre polynomials P_n

$$\begin{cases} V_0 = \dfrac{1 - \cos\theta_c}{2} \\ V_n = \dfrac{V_{sc}}{2}[P_{n-1}(\cos\theta_c) - P_{n+1}(\cos\theta_c)] \text{ for } n>0 \end{cases}.$$
(A3)

The pressure field $p_m(\mathbf{r})$ generated by a monopole located at $\mathbf{r}_0$ between two concentric rigid spheres is given by

$$p_m(\mathbf{r}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} [a_{nm}^o h_n^{(1)}(kr)Y_{nm}(\theta,\phi)$$
$$+ a_{nm}^i j_n(kr)Y_{nm}(\theta,\phi)], \quad \text{for } r<r_0.$$
(A4)

The a_{nm}^o and a_{nm}^i cofficients are given by

$$\begin{cases} a_{nm}^o = \rho_0 c k^2 Q_m \dfrac{h_n^{(1)}(kr_0)j_n'(kb) - h_n'^{(1)}(kb)j_n(kr_0)}{h_n'^{(1)}(kb)j_n'(ka) - h_n'^{(1)}(ka)j_n'(kb)} j_n'(ka)Y_{nm}(\theta_0,\phi_0) \\ a_{nm}^i = \rho_0 c k^2 Q_m \left[h_n^{(1)}(kb)Y_{n-m}(\theta_0,\phi_0) - \dfrac{h_n^{(1)}(kr_0)j_n'(ka) - h_n'^{(1)}(ka)j_n(kr_0)}{h_n'^{(1)}(kb)j_n'(ka) - h_n'^{(1)}(ka)j_n'(kb)} h_n'^{(1)}(kb)Y_{nm}(\theta_0,\phi_0) \right] \end{cases},$$
(A5)

where Q_m is the volumetric velocity of the monopole.

For simulation cases with spherical and secondary monopole sources, solution can be calculated using the superposition theorem.

[1] J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**(4), 1395–1413 (1985).

[2] E. G. Williams, *Fourier Acoustics Sound radiation and Near field Acoustical Holography* (Academic, San Diego, 1999).

[3] R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," in *Proceedings of the ISCV6*, 843–850 (Copenhagen, Denmark, 5–8 July 1999).

[4] E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**(4), 1976–1988 (2001).

[5] Y. I. Bobrovnitskii and T. M. Tomilina, "General properties and fundamental errors of the method of equivalent sources," Acoust. Phys. **41**(5),

649–660 (1995).

[6] M. Ochmann, "The source simulation technique for acoustic radiation problems," Acustica **81**, 512–527 (1995).

[7] Z. Wang and S. F. Wu, "Helmholtz equation - least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**(4), 2020–2032 (1997).

[8] O. Ochmann, "The full-field equation for acoustic radiation and scattering," J. Acoust. Soc. Am. **105**(5), 2574–2584 (1999).

[9] G. Weinreich and E. B. Arnold, "Method for measuring acoustic radiation fields," J. Acoust. Soc. Am. **68**(5), 404–411 (1980).

[10] I. E. Tsukernikov, "Calculation of the field of a found source in a bounded space," Sov. Phys. Acoust. **35**(3), 304–306 (1989).

[11] M. Melon, C. Langrenne, D. Rousseau, B. Roux, and P. Herzog, "Comparison of four subwoofer measurement techniques," *120th AES Convention*, Paris, France, May 20–23 (2006).

[12] A. J. Romano, J. A. Bucaro, B. H. Houston, and E. G. Williams, "On a novel application of the Helmholtz integral in the development of a virtual sonar," J. Acoust. Soc. Am. **108**(6), 2823–2828 (2000).

[13] P. C. Waterman, "New formulation of acoustic scattering," J. Acoust. Soc. Am. **45**(6), 1417–1429 (1969).

[14] Y. I. Bobrovnitskii, K. I. Mal'tser, N. M. Ostapishin, and S. N. Panov, "Acoustical model of a machine," Sov. Phys. Acoust. **37**(6), 570–574 (1991).

[15] I. E. Tsukernikov, "Prediction of radiation field of sound source having locally responsive surface from measurements of diffraction field characteristics," in *Third International Congress on Air—and Structure—Borne Sound and Vibration*, 1727–1730, Montreal (1994).

[16] C. Langrenne, "Méthodes de Régularisation du Probléme Inverse Acoustique Pour L'identification de Sources en Milieu Confiné (Regularization Methods of the Acoustic Inverse Problem for the Characterization of Sources in Bounded Space)," Ph.D. thesis, Maine University, Le Mans, France (1997). In french.

[17] W. Tobocman, "Comparison of the *T*-matrix and Helmholtz integral equation methods for wave scattering calculations," J. Acoust. Soc. Am. **77**(2), 369–374 (1985).

[18] Y. I. Bobrovniskii and G. Pavic, "Modeling and characterization of airborne noise sources," J. Sound Vib. **261**(3), 527–555 (2003).

[19] F. J. Fahy, *Sound Intensity* (Elsevier, London, 1989).

[20] A. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "An advance computational method for radiation and scattering of acoustic waves in three dimensions," J. Acoust. Soc. Am. **77**(2), 362–368 (1985).

[21] H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**(1), 41–58 (1968).

[22] W. Tobocman, "Calculation of acoustic wave scattering by means of the Helmholtz integral equation. I," J. Acoust. Soc. Am. **76**(2), 599–607 (1984).

A mechano-electro-acoustical model for the cochlea: Response to acoustic stimuli

Sripriya Ramamoorthy[a)] and Niranjan V. Deo[b)]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109

Karl Grosh[c)]
Department of Mechanical Engineering and Department of Biomedical Engineering,
University of Michigan, Ann Arbor, Michigan 48109

A linear, physiologically based, three-dimensional finite element model of the cochlea is developed. The model integrates the electrical, acoustic, and mechanical elements of the cochlea. In particular, the model includes interactions between structures in the organ of Corti (OoC), piezoelectric relations for outer hair cell (OHC) motility, hair bundle (HB) conductance that changes with HB deflection, current flow in the cross section and along the different scalae, and the feed-forward effect. The parameters in the model are based on guinea-pig data as far as possible. The model is vetted using a variety of experimental data on basilar membrane motion and data on voltages and currents in the OoC. Model predictions compare well, qualitatively and quantitatively, with experimental data on basilar membrane frequency response, impulse response, frequency glides, and scala tympani voltage. The close match of the model predictions with experimental data demonstrates the validity of the model for simulating cochlear response to acoustic input and for testing hypotheses of cochlear function. Analysis of the model and its results indicates that OHC somatic motility is capable of powering active amplification in the cochlea. At the same time, the model supports a possible synergistic role for HB motility in cochlear amplification.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2713725]

I. INTRODUCTION

In this paper, we develop a linear, physiologically based, three-dimensional finite element model of the cochlea that explicitly couples the electrical, acoustic, and mechanical elements of the cochlea. The normal function of the cochlea relies on a carefully orchestrated tripartite mechanical, electrical, and acoustical (fluidic) coupling. Acoustic stimulation of the cochlea launches a fluid-structure traveling wave along the basilar membrane (BM) and other cochlear structures (e.g., Ref. 1) and a concomitant electrical response (one manifestation of which is the cochlear microphonic, e.g., Ref. 2), each correlated to the input stimulus. Artificial intracochlear electrical stimulation results in emission from the ear[3] and predictably alters the mechanical response of the BM to input sound.[4] Since the discovery of outer hair cell (OHC) somatic electromotility,[5] these cells have been the focus of investigation as a mediator of electrical-structural interaction and of amplification. There exists a body of evidence supporting this hypothesis. For instance, it has been found that drugs that decrease OHC motility *in vitro* likewise decrease the amplitude of the BM velocity and sharpness of the frequency filtering in response to acoustic stimulus (e.g., Ref. 6). OHC electromotility covers the entire mammalian auditory frequency range extending to at least 70 kHz

in vitro[7] and up to 100 kHz *in vivo*.[8] The OHC provides both forward transduction, converting mechanical energy to electrical, and reverse transduction, converting electrical energy to mechanical. Forward transduction also occurs in the OHC stereocilia hair bundle (HB) at acoustic frequencies, as their conductance depends on the rotation of the stereocilia,[9] the so-called mechanoelectrical transducer (MET) sensitivity. The cycle-by-cycle conductance changes are paramount to normal cochlear response as they are hypothesized to drive the OHC electromotility. Reverse HB transduction at acoustic frequencies has been hypothesized[10] as a potential amplification mechanism, leading to the possibility of simultaneous high frequency force generation by the stereocilia and the OHC soma. As measurement techniques are improved and more physiological conditions are used, estimates of the time constant of electrically induced motility of the hair bundle are shifting downward to the tenths of millisecond and postulated to be as fast as 50 μs.[11,12] However, there are no data on electromotile force generation from the stereocilia at frequencies greater than a few Kilohertz.[11,13] Such an effect is not yet included in the model developed in this paper. The coupling of two features, OHC somatic electromotility and HB conductance changes, is included.

Although it has been shown that the OHC electromotility is effective *in vivo*,[8] it is not clear how OHCs can provide amplification at high frequencies in the face of the low transmembrane impedance due to its basolateral membrane capacitance. In order to overcome the reduction in motility due to filtering of the OHC transmembrane potential by the ba-

---

[a)]Currently at Bose Corporation.
[b)]Electronic mail: ndeo@umich.edu
[c)]Electronic mail: grosh@umich.edu

solateral conductance and capacitance, some modelers use an unreasonably large OHC active force[14] or directly compensate for the reduction in OHC active force with frequency.[15] A detailed micromechanically based model is developed in Ref. 16. In that study, the OHC active force is realistic in its magnitude and proportional to OHC hair bundle deflection. However, an additional constant, but still ad hoc, phase factor is added. Lim and Steele[17] develop a purely mechanical model showing that a feed forward model can be used to produce the correct phasing of an active force applied to the BM. However, an analysis of the currents necessary to pass through the active cells is not performed (as the model is purely mechanical) and the levels of the force needed are not provided. In Ref. 18, a realistic OHC active force is employed along with subpartitions [BM, TM, and reticular lamina (RL)] to show nearly 35 dB relative gain between passive and active response. However, the OHC and HB displacement are not explicitly derived from the micromechanics.

We present a method that explicitly models the electrical, fluid, and mechanical domains and their interaction. This key step is circumvented in most other models by using physiologically motivated arguments linking the response of the BM or the stereocilia directly to the force generated by the OHCs. Dimitriadis and Chadwick[19] proposed a fully coupled model, but presented no predictions from it. We present the piezoelectric behavior of the OHC at a more physiological level than has been done previously by explicitly representing the electrical domain and coupling it to the micromechanics. The model matches qualitatively and quantitatively a wide range of measures seen in experiments. Indeed, the purpose of this paper is to illustrate the close match between the experimental results observed at basal locations in the cochlea and the results predicted using the physiologically based mathematical model. Furthermore, we demonstrate robustness of the model predictions to changes in model parameters. The model results show that OHC somatic force generation is quite capable of producing the amplification seen *in vivo*.

II. THE MODEL

The model has been built with an objective to simulate BM response to acoustic signals. With this aim in mind, simplifications were made while creating the cochlear model, in order to focus on the mechanisms involved and to show their interplay. The fundamental components and assumptions can be listed rather readily, but the details of the modeling, naturally, are somewhat involved. In short, the model couples an inviscid, uncoiled, two-duct model for the macroscopic fluid to the BM (see Fig. 1). The locally reacting model of the BM also interacts with a micro-mechanical model of the organ of Corti (OoC) that includes transverse and radial motion of the TM. Viscosity is incorporated through damping of the OoC and the BM. A cable model is used to represent the macroscopic current flow in the scalae. Electromechanical coupling arises from a piezoelectric model for the OHC soma. The AC transducer current through the HBs is driven by the displacement dependent conductance of the HBs which modulates
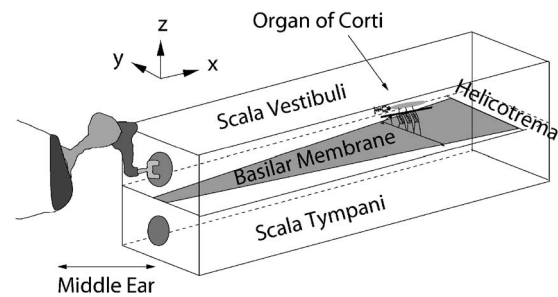
FIG. 1. Illustration of the model used for simulating the cochlea. Organ of Corti is only pictured at one cross section. In the mathematical model it is included over the entire length of the basilar membrane.

the resting potential. We study small perturbations about an operating point. Hence a linearized model is used. These components of the model, their interaction, and the numerical modeling are detailed next.

A. Fluid and the basilar membrane

Figure 1 illustrates a schematic of the standard box model of the cochlea (e.g., Ref. 20) that is used to approximate the fluid domain. The x coordinate is identified with the longitudinal direction of the cochlea, while the y and z coordinates represent a cross section of the cochlea with the z coordinate normal to the BM (the transverse direction), and the y coordinate in the radial direction. The box model for the cochlea is assumed to have rigid walls and is filled with fluid similar to water. The macroscopic fluid response is assumed to be incompressible and nonviscous. For the purpose of this study only harmonic motions of the system were considered. A time dependence of $e^{i\omega t}$ is assumed, where ω is the angular frequency. The governing equation for the (incompressible, inviscid) fluid is

$$\nabla^2 p = 0, \tag{1}$$

where p is the pressure in the fluid. The duct is divided into two equal sections by the BM which lies in the $x-y$ plane. The BM extends to the helicotrema, which is modeled as a 1 mm hole connecting the two ducts. For computational efficiency a modal decomposition of the fluid is used in the radial (y) direction, as in Ref. 20,

$$p(x,y,z) = \sum_{m=0}^{M} p_m(x,z)\cos\left(\frac{m\pi(y + w/2)}{w}\right),$$

$$-\frac{w}{2} \le y \le \frac{w}{2}, \tag{2}$$

where w is the width of the duct, m is the mode number, and M is the total number of modes used in the y direction. The decomposition results in a series of M two-dimensional problems to solve for $p_m(x,z)$. The fluid velocity is related to the fluid pressure by the linearized Euler relation,

$$\nabla p = -i\rho\omega\mathbf{v_f}, \tag{3}$$

where ρ is the density of the fluid (water), and $\mathbf{v_f}$ is the fluid velocity.

The BM is modeled as a set of parallel simply supported beams. Hence, longitudinal stiffness coupling is not included

in the model, as we attribute to the fluid the main coupling in the x direction. Arguments for including longitudinal structural coupling in the BM have been made previously (e.g., Ref. 21). The present model can be extended to include such effects but we have not done so in this paper. The BM motion is decomposed into a sum of modes in the radial direction,

$$u_{\mathrm{bm}}(x,y) = \sum_{n=1}^{N} u_{\mathrm{bm}}^{(n)}(x)\Psi_n(y), \qquad (4)$$

where

$$\Psi_n(y) = \sin\left(\frac{n\pi(y+b/2)}{b}\right), \quad -b/2 \leqslant y \leqslant b/2,$$

and b is the width of the membrane. In this study we consider multiple modes for the fluid but only the first mode for the BM. Higher modes of the BM are not included because motion of the BM seen *in vivo* is quite similar to the first symmetric mode.[22] In deriving the equations of motion (Sec. II C) and kinematic relations (Sec. II B 2) only the first mode of the BM is considered. The superscript for the first BM mode has been dropped in subsequent equations in this paper.

Taking advantage of the orthogonality of modes, a series of two-dimensional equations can be obtained for the fluid and the BM by integrating out the radial dependence:

$$\frac{\partial^2 p_m}{\partial x^2} + \frac{\partial^2 p_m}{\partial z^2} - \left(\frac{m\pi}{w}\right)^2 p_m(x,z) = 0, \quad 0 \leqslant m \leqslant M, \qquad (5)$$

$$\frac{b}{2}(-M_{\mathrm{bm}}(x)\omega^2 + iC_{\mathrm{bm}}(x)\omega + K_{\mathrm{bm}}(x))u_{\mathrm{bm}}(x)$$

$$= \sum_{m=1}^{M} (p_m^{\mathrm{SV}} - p_m^{\mathrm{ST}})\mu_m + Q_{\mu\mathrm{mech}}, \qquad (6)$$

where M_{bm}, C_{bm}, and K_{bm} are the resultant mass, damping, and stiffness, respectively, for the first structural mode; p_m^{SV} and p_m^{ST} represent the pressure loading on the BM from the fluid corresponding to mode m in the scala vestibuli and scala tympani, respectively; μ_m is the coupling coefficient resulting from the integration of the lateral modes for pressure and BM displacement,[20]

$$\mu_m = \int_{-b/2}^{b/2} \cos\left(\frac{m\pi(y+w/2)}{w}\right)\sin\left(\frac{\pi(y+b/2)}{b}\right)dy, \qquad (7)$$

and $Q_{\mu\mathrm{mech}}$ is the force from the micro-electro-mechanical model described in Sec. II C, Eq. (32). The factor $b/2$ in Eq. (6) arises from the integration of the first BM mode over the width of the BM.

B. Micro-electro-mechanical model

1. Electrical environment

The macroscopic spread of current through the different scalae is modeled using one-dimensional cables.[23] Figure 2 shows the circuit at a given cross section in the cochlea. The model includes cables for current flow in the scala vestibuli (SV), the scala tympani (ST), and the scala media (SM). The
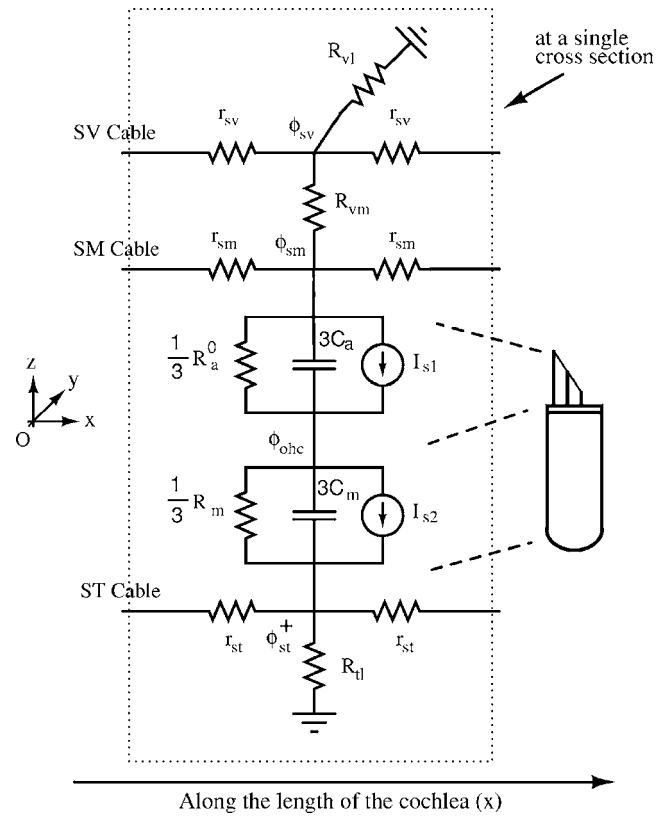
FIG. 2. Electrical network at a given cross section of the cochlea. The SV, SM, and ST cables run longitudinally along the cochlea, ϕ_{st}^+ represents ST potential at an apical location $(x+\delta x)$ in the cochlea relative to the other quantities. This allows for the possibility of forward inclination of the OHCs (see the feed-forward effect in Sec. II B 4). The factor of three multiplying the apical and the basolateral membrane capacitance, and dividing the apical and the basolateral membrane resistance of the OHC is to account for the three OHCs in a cross section.

ST cable represents the potential in the ST very close to the BM or just in the interstitial space of the OoC, outside the OHCs. Away from the BM, the ST is modeled as being nearly ground. Resistances r_{sv}, r_{sm}, r_{st} represent resistance per unit length along the SV, SM, and ST cables, respectively; R_{vm} is the resistance seen by the current flowing from the SV to the SM; R_{vl} is the resistance to current flowing from SV to ground; R_{tl} is the resistance to flow of current flowing from ST/interstitial space to ground. R_a^0 and C_a represent the apical resistance and capacitance, while R_m and C_m represent the basolateral resistance and capacitance, respectively, of the outer hair cells. I_{s1} and I_{s2} are current sources due to the variable HB conductance and OHC electromotility, respectively [explained in Sec. II B 3, see Eqs. (20) and (23)]. The voltages shown in Fig. 2 represent the fluctuating (or AC) part of voltages in the different scalae. The voltages are at the same cross-section except for the ST voltage. Due to the forward inclination of the OHCs (the feed-forward effect), the ST voltage is located at a certain distance (the feed-forward distance) apical to the location of the other three voltages [see Sec. II B 4, and Eq. (28)].

The electrical domain equations are determined using Kirchhoff's laws. In each cross-sectional circuit branch, there are four electrical potentials, the SV (ϕ_{sv}), the SM (ϕ_{sm}), the OHC (ϕ_{ohc}), and the ST (ϕ_{st}). In addition to cur-
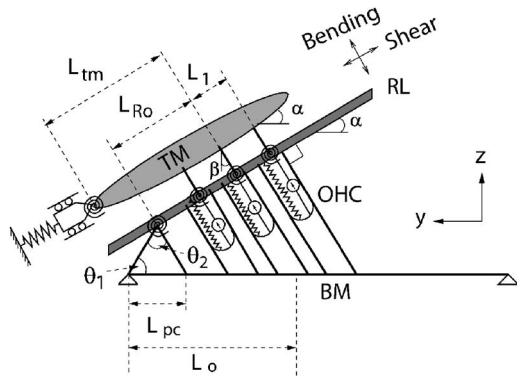
FIG. 3. Micromechanical model for the organ of Corti structures. The sketch is drawn for $\alpha = \beta$ and is pictured here with no feed-forward in the x direction. BM: Basilar membrane; TM: Tectorial membrane; OHC: Outer hair cell; RL-Reticular Lamina. L_{Ro} is the distance between RL pivot point on the arch of Corti and the middle OHC; L_1 is the radial distance between OHCs along the RL; θ_1 is the acute angle between the inner pillar cell and the BM, θ_2 is the acute angle between the inner and outer pillar cell, L_{pc} is the distance along the BM between the left edge of the BM and the contact point of the outer pillar cell with the BM, L_o is the distance along the BM between the left edge of the BM and the contact point of the middle OHC with the BM, and L_{tm} is the length of the TM from its pivot to the attachment point of the middle row HBs with the TM.

rent flow in the cross section, current flow along the length of the cochlea is also allowed in the three scalae. Applying Kirchhoff's laws (in both longitudinal and cross-sectional dimensions) to the circuit shown in Fig. 2 yields

$$\frac{1}{r_{sv}}\frac{\partial^2 \phi_{sv}}{\partial x^2} - \left(\frac{1}{R_{vl}} + \frac{1}{R_{vm}}\right)\phi_{sv} + \frac{1}{R_{vm}}\phi_{sm} = 0, \qquad (8)$$

$$\frac{1}{R_{vm}}\phi_{sv} + \frac{1}{r_{sm}}\frac{\partial^2 \phi_{sm}}{\partial x^2} - \left(\frac{1}{R_{vm}} + 3Y_a\right)\phi_{sm}$$
$$+ 3Y_a \phi_{ohc} - I_{s1} = 0, \qquad (9)$$

$$3Y_a \phi_{sm} - 3(Y_a + Y_m)\phi_{ohc} + 3Y_m \phi_{st}^+ + I_{s1} - I_{s2} = 0, \quad (10)$$

$$3Y_m \phi_{ohc} + \frac{1}{r_{st}}\frac{\partial^2 \phi_{st}^+}{\partial x^2} - \left(\frac{1}{R_{tl}} + 3Y_m\right)\phi_{st}^+ + I_{s2} = 0. \qquad (11)$$

In Eqs. (8)–(11), $Y_a = 1/R_a^0 + i\omega C_a = G_a^0 + i\omega C_a$ and $Y_m = 1/R_m + i\omega C_m$ are the admittances of the apical and basolateral portions of the OHC, respectively. The + superscript indicates a location apical to the longitudinal location under consideration as explained in the section on feed-forward (Sec. II B 4). In this model, the only current path from the SM to the ST is assumed to exist at the apical pole of the OHC and all other junctions at this interface are assumed to be tight.

2. Kinematic model of microstructures

Figure 3 shows a representation of the kinematic model of the (OoC) used in the model. A kinematic model is constructed based on Ref. 24 with different stiffnesses being represented as springs at various locations. The masses of the organ of Corti structures are lumped onto the TM and the BM. The TM is assumed to have two degrees of freedom—a radial (or shear) mode and a transverse (or bending) mode (see Fig. 3). Damping in the system is accounted for through

structural damping of the BM and the viscous damping associated with TM motion. Coupling arising from phalangeal processes is neglected, but the feed-forward effect arising from OHC inclination is included. Since the Deiters cell (DC) impedance is in series with OHC impedance and is at least a couple of orders higher in magnitude,[21] the DC is taken as a rigid connection. The RL is modeled as a massless rigid bar. The HBs are also modeled as rigid links. Note that in this representation, the RL and TM are not constrained to stay parallel and the fluid in the sub-tectorial space is not modeled explicitly. The effects of the pressure and fluid motion associated with squeezing the subtectorial fluid layer are not yet included in this model. Only the BM interacts directly with the fluid.

The equations of motion for the OoC are expressed in terms of the BM and the two TM degrees of freedom using Lagrange's method (see Sec. II C). In deriving the equations of motion, the displacement of the OHCs, and the rotation of the HBs and the RL are required. In this section, we list the kinematic assumptions used to express motion of the OHC, HB, and RL in terms of the BM and the TM motion. The kinematic equations obtained in the following are for very small motions. The OHC and HB angles (α and β, respectively) referred to in Fig. 3 pertain to only the middle row OHCs and HBs. Kinematic relations for the middle row OHCs and HBs are first derived ignoring the rigid HB links from the first and third row OHCs (otherwise it is a locked mechanism). The kinematic relations for the first and third row cells are then extrapolated from the middle row relations.

The arch of Corti is assumed to be rigid and hence rotates when the BM is displaced. The displacement of the apical end of the arch of Corti (top of the pillar cells) is therefore given by

$$u_{ap}(x) = u_{bm}(x)\Psi_1(b/2 - L_{pc})\frac{\sin(\theta_1 + \theta_2)}{\sin(\theta_2)}, \qquad (12)$$

where $u_{bm}(x)$ is the amplitude of the first mode of BM displacement at location x and $\Psi_1(b/2 - L_{pc})$ is the first mode shape of the BM evaluated at $y = b/2 - L_{pc}$ (b is the width of the BM). To determine OHC compression, deflections of the OHC-apex and OHC-base along the OHC are computed. The motion of the middle row OHC-apex toward the BM and along the OHC is given by

$$u_{ohc_2}^a(x) = u_{bm}(x)\Psi_1(b/2 - L_{pc})\frac{\sin(\theta_1 + \theta_2)}{\sin(\theta_2)}$$
$$\times \left(-\cos(\theta_1 - \alpha) + \frac{\cos(\theta_1 - \beta)}{\cos(\alpha - \beta)}\right)$$
$$- u_{tms}(x)\tan(\alpha - \beta) - u_{tmb}(x). \qquad (13)$$

The motion of the middle row OHC-base toward the RL and along the OHC is

$$u_{ohc_2}^b(x) = u_{bm}(x)\Psi_1(b/2 - L_o)\cos(\alpha). \qquad (14)$$

The total compression of the OHC will be the sum of these displacements. However when feed-forward is present, the apical end displacement and basal end displacement of the

OHC will occur at different longitudinal (x) locations. The details of computing the OHC deformation associated with a feed-forward model are discussed in Sec. II B 4.

The twist in the coil spring attached to the HBs is a sum of the rotation of the HBs (due to BM and TM motion) and rotation of the RL (due to BM motion). For small motions, the relative shear motion between the top end (the end connected to the TM) of the middle row HB and its bottom end (connected to the RL), in a direction normal to the HB and toward the outer rows of HBs (toward negative y direction in Fig. 3) is

$$u_{\mathrm{hb}_2}(x) = u_{\mathrm{bm}}(x)\Psi_1(b/2 - L_{\mathrm{pc}})\frac{\sin(\theta_1 + \theta_2)}{\sin(\theta_2)}\frac{\sin(\theta_1 - \alpha)}{\cos(\alpha - \beta)}$$
$$+ \frac{u_{\mathrm{tms}}(x)}{\cos(\alpha - \beta)}. \tag{15}$$

This HB shear divided by the height of the HB yields the HB rotation relative to its resting position. For small motions, the RL displacement at the location of the middle OHC relative to its (the RLs) attachment at the pillar cells, in a direction normal to itself (RL) and away from the OHC (toward positive z direction in Fig. 3) is given by

$$u_{\mathrm{rl}_2}(x) = - u_{\mathrm{bm}}(x)\Psi_1(b/2 - L_{\mathrm{pc}})\frac{\sin(\theta_1 + \theta_2)}{\sin(\theta_2)}\frac{\cos(\theta_1 - \beta)}{\cos(\alpha - \beta)}$$
$$+ u_{\mathrm{tms}}(x)\tan(\alpha - \beta) + u_{\mathrm{tmb}}(x). \tag{16}$$

This RL motion at the middle OHC normal to itself divided by distance from the pivot (L_{ro} for the middle OHC) gives the rotation of the RL relative to its resting position. The sum of the HB rotation and RL rotation yields the angle of twist in the coil spring connecting the HBs to the RL.

These quantities can then be derived for the first and third row OHCs and HBs using the middle row displacements as reference and the relative lever ratios. For example, the displacement of the apical end of the first row OHC can be written as

$$u_{\mathrm{ohc}_1}^a(x) = u_{\mathrm{bm}}(x)\Psi_1(b/2 - L_{\mathrm{pc}})\frac{\sin(\theta_1 + \theta_2)}{\sin(\theta_2)}$$
$$\times\left(-\cos(\theta_1 - \alpha) + \left[1 - \frac{L_1}{L_{\mathrm{Ro}}}\right]\frac{\cos(\theta_1 - \beta)}{\cos(\alpha - \beta)}\right)$$
$$-\left(1 - \frac{L_1}{L_{\mathrm{Ro}}}\right)u_{\mathrm{tmb}}(x) - \left(1 - \frac{L_1}{L_{\mathrm{Ro}}}\right)$$
$$\times u_{\mathrm{tms}}(x)\tan(\alpha - \beta). \tag{17}$$

These corrections for the first and third row locations are small and can be neglected. They were however included for the results presented in this paper.

3. Hair bundle conductance and OHC electromotility

The coupling between the electrical domain and the mechanical domain occurs through the variable conductance of the HBs and electromotility of the OHCs. The HBs are assumed to have a conductance which changes linearly with the deflection of the HB,

$$G_{a_j} = G_a^0 + G_a^1 u_{\mathrm{hb}_j}. \tag{18}$$

Here G_{a_j} is the conductance of the jth HB as a function of its deflection u_{hb_j}, where $j = 1, 2, 3$ is the radial counting index for the OHC. G_a^0 represents conductance at the resting state ($G_a^0 = 1/R_a^0$) of the HB and G_a^1 represents slope of the change of conductance with respect to the HB deflection; we will denote this as the mechanoelectrical transducer (MET) sensitivity. G_a^0 and G_a^1 are assumed constant at a given cross section. The current flowing through the jth HB (I_{hb_j}) is the product of the apical admittance and the potential drop from the SM to the OHC interior. The linearized expression for this current is

$$I_{\mathrm{hb}_j} = (G_a^0 + i\omega C_a)(\phi_{\mathrm{sm}} - \phi_{\mathrm{ohc}}) + (V_{\mathrm{sm}} - V_{\mathrm{ohc}})G_a^1 u_{\mathrm{hb}_j}, \tag{19}$$

where V_{sm} and V_{ohc} are the voltages at resting state in the SM and the OHC, respectively, and C_a is the apical capacitance of the OHC. As explained in Sec. II B 1, the quantities ϕ_{sm} and ϕ_{ohc} are the fluctuating parts of the voltages in the SM and the OHC, respectively, and are assumed to be the same for each HB at a given cross section. The contribution of the variable conductance to the total HB current at a given cross section is expressed as an equivalent current source I_{s1} in the model (see Fig. 2) whose value is given by

$$I_{s1} = (V_{\mathrm{sm}} - V_{\mathrm{ohc}})G_a^1\sum_{j=1}^{3} u_{\mathrm{hb}_j}. \tag{20}$$

The OHC electromotility is modeled through linearized expressions relating OHC strain and transmembrane voltage to the OHC force and current[25] as

$$F_{\mathrm{ohc}_j} = K_{\mathrm{ohc}}u_{\mathrm{ohc}_j}^{\mathrm{comp}} + \epsilon_3(\phi_{\mathrm{ohc}} - \phi_{\mathrm{st}}^+), \tag{21}$$

$$I_{\mathrm{ohc}_j} = (\phi_{\mathrm{ohc}} - \phi_{\mathrm{st}}^+)/Z_m - i\omega\epsilon_3 u_{\mathrm{ohc}_j}^{\mathrm{comp}}. \tag{22}$$

Here F_{ohc_j} is the force exerted by the jth OHC on the BM and the RL, I_{ohc_j} is the current flowing through the jth OHC, $u_{\mathrm{ohc}_j}^{\mathrm{comp}}$ represents compression in the jth OHC, ϵ_3 is the electromechanical coupling coefficient, Z_m is the net basolateral impedance of the OHC given by $1/Z_m = 1/R_m + i\omega C_m$ (see Fig. 2), $(\phi_{\mathrm{ohc}} - \phi_{\mathrm{st}}^+)$ is the alternating (AC) basolateral transmembrane potential (the term ϕ_{st}^+ is explained in Sec. II B 4, and K_{ohc} is the stiffness of the OHC. The potentials and OHC properties ($K_{\mathrm{ohc}}, \epsilon_3, Z_m$) are assumed to be the same for each OHC at a given cross section. The current source I_{s2} shown in Fig. 2 corresponds to the total current due to the piezoelectric-like behavior of the OHC,

$$I_{s2} = - i\omega\epsilon_3\sum_{j=1}^{3} u_{\mathrm{ohc}_j}^{\mathrm{comp}}. \tag{23}$$

Equations (19) and (22) show how deformation of the organ of Corti leads to additional current flow through the

Ramamoorthy *et al.*: Modeling acoustic response of the cochlea

HBs and the OHCs, while Eq. (21) shows how modulation of the OHC transmembrane potential leads to an equivalent force generation by the OHCs.

4. Feed-forward

The OHCs *in vivo* are inclined toward the base of the cochlea. This inclination is hypothesized to "feed-forward" the energy[26] in the cochlea. Therefore, the force exerted by an OHC is not completely in the $y-z$ plane. To include the feed-forward effect, we now have to decompose F_{ohc_j} into apical and basal locations.

$F^a_{\mathrm{ohc}_j}$ represents force at the apical end (or top end) of the jth OHC. Similarly $F^b_{\mathrm{ohc}_j}$ represents force at the basal end (or bottom end) of the jth OHC. $F^a_{\mathrm{ohc}_j}(x)$ depends on field quantities at x and at $x+\delta x$ while $F^b_{\mathrm{ohc}_j}(x)$ depends on field quantities at x and $x-\delta x$, where δx is the feed-forward distance,

$$F^a_{\mathrm{ohc}_j}(x) = (K_{\mathrm{ohc}}(u^a_{\mathrm{ohc}_j} + u^{b+}_{\mathrm{ohc}_j}) + \epsilon_3(\phi_{\mathrm{ohc}} - \phi^+_{\mathrm{st}}))\cos(\psi),$$
(24)

$$F^b_{\mathrm{ohc}_j}(x) = (K_{\mathrm{ohc}}(u^{a-}_{\mathrm{ohc}_j} + u^b_{\mathrm{ohc}_j}) + \epsilon_3(\phi^-_{\mathrm{ohc}} - \phi_{\mathrm{st}}))\cos(\psi),$$
(25)

where ψ is the angle of forward inclination (toward the base) of the OHCs with the vertical. Expressions for $u^a_{\mathrm{ohc}_j}$ and $u^b_{\mathrm{ohc}_j}$ are given in Eqs. (13) and (14) in Sec. II B 2. The + and − superscripts represent locations apical or basal, respectively, to the given location in the cochlea, i.e., the displacements and voltages at location $x+\delta x$ and $x-\delta x$, respectively, if x is the location under consideration. Since the feed-forward distance is small ($\approx 5-10\ \mu\mathrm{m}$), in our implementation we use the Taylor series approach and retain only the first-order terms. That is,

$$u^{b+}_{\mathrm{ohc}_j} = u^b_{\mathrm{ohc}_j}(x + \delta x) = u^b_{\mathrm{ohc}_j}(x) + \frac{du^b_{\mathrm{ohc}_j}(x)}{dx}\delta x,$$
(26)

$$u^{a-}_{\mathrm{ohc}_j} = u^a_{\mathrm{ohc}_j}(x - \delta x) = u^a_{\mathrm{ohc}_j}(x) - \frac{du^a_{\mathrm{ohc}_j}(x)}{dx}\delta x,$$
(27)

$$\phi^+_{\mathrm{st}} = \phi_{\mathrm{st}}(x + \delta x) = \phi_{\mathrm{st}}(x) + \frac{d\phi_{\mathrm{st}}(x)}{dx}\delta x,$$
(28)

$$\phi^-_{\mathrm{ohc}} = \phi_{\mathrm{ohc}}(x - \delta x) = \phi_{\mathrm{ohc}}(x) - \frac{d\phi_{\mathrm{ohc}}(x)}{dx}\delta x.$$
(29)

C. Local governing equations

The derivation of the local equations for the mechanical degrees of freedom is described in this section. The dependence of the variables on the axial location x has been assumed and not explicitly written. The kinetic and potential energy for this whole system, at a given location x, can then be written as

$$\mathcal{T} = \frac{1}{2}M_{\mathrm{tms}}\dot{u}^2_{\mathrm{tms}} + \frac{1}{2}M_{\mathrm{tmb}}\dot{u}^2_{\mathrm{tmb}} + \frac{1}{2}M_{\mathrm{bm}}\dot{u}^2_{\mathrm{bm}},$$
(30)

$$\mathcal{V} = \frac{1}{2}K_{\mathrm{bm}}u^2_{\mathrm{bm}} + \frac{1}{2}K_{\mathrm{tms}}u^2_{\mathrm{tms}} + \frac{1}{2}K_{\mathrm{tmb}}u^2_{\mathrm{tmb}} + \frac{1}{2}K_{\mathrm{hb}}$$

$$\times \left(\left(u_{\mathrm{hb}_1} + u_{\mathrm{rl}_1}\frac{L_{\mathrm{hb}}}{L_{\mathrm{Ro}} - L1} \right)^2 + \left(u_{\mathrm{hb}_2} + u_{\mathrm{rl}_2}\frac{L_{\mathrm{hb}}}{L_{\mathrm{Ro}}} \right)^2 \right.$$

$$\left. + \left(u_{\mathrm{hb}_3} + u_{\mathrm{rl}_3}\frac{L_{\mathrm{hb}}}{L_{\mathrm{Ro}} + L_1} \right)^2 \right) + \frac{1}{2}K_{\mathrm{rl}}\left(u_{\mathrm{rl}_2} + \frac{u_{\mathrm{ap}}}{L_{\mathrm{po}}}L_{\mathrm{Ro}} \right)^2,$$
(31)

where u_{tms}, K_{tms}, M_{tms} represent the displacement, stiffness, and mass of the TM shear mode, respectively, and u_{tmb}, K_{tmb}, M_{tmb} represent corresponding quantities for the TM bending mode. In Eqs. (30) and (31) the rotational stiffnesses of the coil springs for the HBs, the TM, and the RL are converted to equivalent linear stiffnesses after multiplying with appropriate lengths.

The generalized work done by external forces (including the electromotile forces from the OHCs, the acoustic pressure, and nonconservative viscous damping) is given by

$$Q = -\sum_{j=1}^3 F^a_{\mathrm{ohc}_j} u^a_{\mathrm{ohc}_j}$$

$$-\sum_{j=1}^3 F^b_{\mathrm{ohc}_j} u^b_{\mathrm{ohc}_j} - \sum_m (p^{\mathrm{SV}}_m - p^{\mathrm{ST}}_m)u_{\mathrm{bm}}\mu_m$$

$$- (C_{\mathrm{bm}}\dot{u}_{\mathrm{bm}})u_{\mathrm{bm}} - (C_{\mathrm{tms}}\dot{u}_{\mathrm{tms}})u_{\mathrm{tms}} - (C_{\mathrm{tmb}}\dot{u}_{\mathrm{tmb}})u_{\mathrm{tmb}}.$$
(32)

The net normal fluid force acting on the TM from the sulcus and SM is assumed to be smaller than the normal HB forces acting on the TM. Hence, the TM interaction with the fluid is not included. The BM interacts with the fluid through the pressure difference in the SV and ST [see Eq. (6)]. The expression for $F^a_{\mathrm{ohc}_j}$ and $F^b_{\mathrm{ohc}_j}$ can be obtained from Eqs. (24) and (25). The terms C_{bm}, C_{tms}, and C_{tmb} represent viscous damping coefficients for BM, TM bending, and TM shear modes, respectively.

The governing equations of motion for the mechanical variables, including coupling to the electrical and acoustic domains, are found from the variation of the Lagrangian ($L = \mathcal{T} - \mathcal{V}$) with respect to the three mechanical variables (u_{bm}, u_{tms}, and u_{tmb}):

$$\frac{\partial}{\partial t}\frac{\partial L}{\partial \dot{u}_j} - \frac{\partial L}{\partial u_i} = \frac{\partial Q}{\partial u_i},$$
(33)

where i varies over bm, tms, and tmb. The expressions for the final set of equations are not listed due to space constraints.

D. Finite elements

The expressions given in the preceding sections constitute the strong form of the equations. These are used to derive the weak form after multiplying with corresponding weighting functions and integrating over the domain.[20] The final set of equations have the following form:

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Ramamoorthy *et al.*: Modeling acoustic response of the cochlea 2763

TABLE I. Geometric data for the macromodel (see Fig. 3).

Property	Value	Source
BM width (b)	80 μm (base) to 200 μm (apex)	Ref. 28
BM thickness (h)	7 μm (base) to 1 μm (apex)	Ref. 28
L_{pc}	$b/3$	Adapted from Refs. 33 and 24
L_1	9 μm	Based on OHC diameter
L_0	$b/2$	Adapted from Refs. 33 and 24
L_{tm}	60 μm (base) to 180 μm (apex)	Adapted from Refs. 33 and 24
L_{Ro}	Calculated from Fig. 3	
α	25° (base) to 45° (apex)	Ref. 34
β	25° (base) to 45° (apex)	Assumed
θ_1	60°	Adapted from Refs. 33 and 24
θ_2	60°	Adapted from Refs. 33 and 24
L_{ohc}	20 μm (base) - 85 μm (apex)	Ref. 34
L_{hb}	1 μm (base) - 6 μm (apex)	Ref. 9
Duct height for SV, ST (H)	1 mm	Ref. 28
Duct width (W)	1 mm	Ref. 28
Duct length (L)	25 mm	Ref. 28

$$\begin{bmatrix} \mathbf{K_f} & \mathbf{Q_{fs}} & \mathbf{0} \\ \mathbf{Q_{sf}} & \mathbf{K_s} & \mathbf{Q_{se}} \\ \mathbf{0} & \mathbf{Q_{es}} & \mathbf{K_e} \end{bmatrix} \begin{pmatrix} \mathbf{p} \\ \mathbf{u} \\ \phi \end{pmatrix} = \begin{pmatrix} \mathbf{f_f} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \tag{34}$$

The matrix is unsymmetric due to the nonreciprocal MET function ($\mathbf{Q_{es}} \neq \mathbf{Q_{se}^T}$) and the feed-forward effect. Linear spatial interpolations are used for the structural and electrical degrees of freedom while bilinear interpolations are used for the fluid. The quantity $\mathbf{K_f}$ is the dynamic stiffness of the fluid, $\mathbf{K_s}$ is the dynamic stiffness of the micromechanical structures, and $\mathbf{K_e}$ represents the interaction among the electrical degrees of freedom. Among the off-diagonal terms, $\mathbf{Q_{es}}$ and $\mathbf{Q_{se}}$ represent electrical-structural coupling at the OHCs, and $\mathbf{Q_{fs}}$ and $\mathbf{Q_{sf}}$ are from fluid-structure coupling at the BM. On the right-hand side, $\mathbf{f_f}$ represents any forcing on the fluid. Using a reference displacement at the stapes, the linear matrix equation is solved to determine the fluid pressure ($\mathbf{p}$), structural displacements ($\mathbf{u}$), and electrical potentials (ϕ) at the nodes.

III. MODEL PARAMETERS

The model has a necessarily large number of parameters. The values are listed in Tables I–III. Where possible, we used or extrapolated parameters from published data (references indicated in the tables). In this section, we describe how the parameters were selected and the assumptions underlying those data for which hard experimental values are not available.

A. Mechanical parameters

The effective BM stiffness for the first mode is calculated from the radial flexural rigidity estimated in Ref. 27. For a simply supported BM, the radial flexural rigidity (D) is estimated to be about 2.5×10^{-1} N m for basal locations in the guinea pig.[27] Using this value, one can compute the effective BM stiffness for the first mode,

$$K_{bm} = D \left(\frac{\pi}{b} \right)^4, \tag{35}$$

where b is the width of the BM. The BM thickness (h) and width (b) are estimated from Ref. 28. Although variation in D along the length of the BM was not found to be statistically significant,[27] the radial flexural rigidity is expected to depend on the BM thickness h ($D \propto h^3$). We therefore include a cubic dependence of D on the BM thickness h. Accounting for this difference in interpretation of D, and estimating the BM thickness to be 6.6 μm and BM width to be 90 μm at about 1 to 2 mm from the base (corresponding to the location where point stiffness measurements were made by Gummer et al.[27]), the estimated K_{bm} for the most basal location ($x=0$) works out to be 6.69×10^7 N/m^2. We use a value of 6×10^7 N/m^2 at $x=0$ in the model.

HB stiffness measurements were made in guinea pigs by Strelioff and Flock.[29] Those measurements were extrapolated to indicate average stiffness of HBs ranging from 63.9 to 284 mN/m for different rows of OHCs at the extreme base of the cochlea ($x=0$). To choose a value for the model the following considerations were made. Using a density of 100

TABLE II. Mechanical properties for the macromodel. x is in meters.

Property	Value	Source
BM stiffness (K_{bm})	$6 \times 10^5 (h/h_0)^3 (b_0/b)^4$ N/m³	Ref. 27
TM bending stiffness (K_{tmb})	$6 \times 10^3 e^{(-220x)}$ N/m²	Ref. 31
TM shear stiffness (K_{tms})	$6 \times 10^3 e^{(-220x)}$ N/m²	Ref. 31
RL stiffness (K_{rl})	$7.6 \times 10^3 e^{(-325x)}$ N/m²	Based on Ref. 24
HB stiffness (K_{hb})	$3.42 \times 10^4 e^{(-325x)}$ N/m²	Estimated from Ref. 29
OHC stiffness (K_{ohc})	$7.6 \times 10^3 e^{(-325x)}$ N/m²	Ref. 30 (see Sec. III A)
BM mass per unit area (M_{tm})	$\rho_{bm} h$ kg/m²	Density assumed
	$\rho_{bm} = 1000$ kg/m³	
BM viscous damping (C_{bm}) (includes fluid viscosity)	0.05 N s/m²	Assumed
TM bending damping (C_{tmb})	0.05 N s/m²	Assumed
TM shear damping (C_{tms})	0.03 N s/m²	Assumed
Effective TM shear mass (M_{tms})	$\rho_{tm} h_{tm}^0 b_{tm}^0 e^{50x}$ kg/m	x variation assumed
	$\rho_{tm} = 1000$ kg/m³, $h_{tm}^0 = 18$ μm, $b_{tm}^0 = 60$ μm	
Effective TM bending mass (M_{tmb})	$0.7 M_{tms}$ kg/m	Assumed
Round window stiffness	1.8×10^5 N/m³	Assumed
Round window damping	5.8×10^4 N s/m³	Assumed

rows of HBs per mm for guinea pigs[9] yields a stiffness ranging from 6.39 to 28.4 kN/m² at $x=0$. Strelioff and Flock[29] used a large displacement of 1 μm to measure stiffness of HBs and their measurements were not taken instantaneously at the application of stimulus. A recent study by Kennedy et al.[11] has shown that under excitatory stimulation, the HB

TABLE III. Electrical properties for the macromodel. x in meters.

Property	Value	Source
ϵ_3	$(-8 \times 10^{-6} - 8 \times 10^{-5}x)$ N/m/mV	Ref. 38
$1/R_a^0$	100 μS/m	Estimated using Ref. 36
$1/R_m$	5100 μS/m (base) - 360 μS/m (apex)	Ref. 35
C_a	50 nF/m	Estimated using Ref. 36
C_m	1800 nF/m (base) - 4200 nF/m (apex)	Ref. 35
G_a^1	variable (see Sec. III B)	Free parameter
$V_{sm} - V_{ohc}$	$(150-1000x)$ mV	Based on Ref. 37
R_{vl}	10 Ωm	Based on Ref. 37
R_{tl}	4 Ωm	Based on Ref. 37
R_{vm}	25 Ωm	Based on Ref. 37
r_{sv}	3 MΩ/m	Based on Ref. 37
r_{sm}	5 MΩ/m	Based on Ref. 37
r_{st}	0.15 Ω/m	Based on R_{OC} in Ref. 37

stiffness noticeably reduces with time. Because of the latter point, at $x=0$ a slightly higher number of 34 kN/m² is used as the average stiffness for HBs in the model. Strelioff and Flock[29] further showed that the HB stiffness decreased at an exponential rate from the base. The decay rate for the HB stiffness varied from -238 to -406 m^{-1} between different rows of OHCs. For simplicity a value of -325 m^{-1} for the decay rate is chosen, which is near the middle of the published range.

OHC stiffness values are available in literature[30]. The stiffness at resting state *in vivo* and its variation along the length of the cochlea are not known exactly. We use a resting stiffness value of about 11 mN/m for OHCs located 6 mm from the base. The value listed in Table II gives OHC stiffness per meter assuming 100 rows of OHCs in 1 mm.[9] The spatial variation of OHC stiffness along the length of the BM is assumed to be same as that of the HB stiffness.

The stiffnesses of other microstructures (TM, RL) are difficult to estimate, in part because the mode of operation (kinetics and kinematics) is still not completely known. In this paper, we put forth a hypothesis of the dominant kinematics that follows closely that proposed by Dallos.[24] We use estimates of the TM stiffness from Zwislocki and Cefaratti[31] as a guide to choose TM bending and shearing stiffness values. The mass of the TM is calculated from dimensions of the TM and using the density of water (1000 kg/m³). Following Gummer et al.,[32] the TM bending mass is chosen to be 30% lower than TM shear mass (results indicate that TM bending resonance is higher than TM shear resonance[32]).
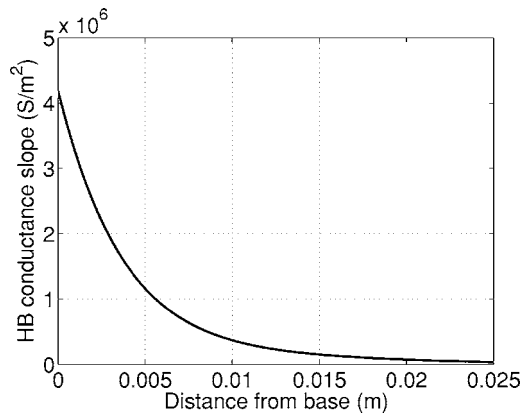
FIG. 4. The plot vs x of the function [Eq. (36)] used for the MET channel sensitivity in the model.

FIG. 6. BM phase at different locations in cycles (2π rad). Dash curves: Passive $(G_a^1(0)=0$ S/m$^2)$ and solid curves: Active $(G_a^1(0)=4.19 \times 10^6$ S/m$^2)$. Locations are (from left to right) 11.75, 10.75, 9.75, 8.75, 7.75, 6.75, and 5.75 mm.

Based on the calculations in Dallos,[24] the value for the RL stiffness [K_{rl}, see Eq. (31)] is chosen to be similar to that of the OHC stiffness. The spatial variation of the RL stiffness is not known and is taken to be the same as that of the HB and OHC stiffness. The viscous modal damping for the three structural degrees of freedom is also not known. In the model, the structural damping is supposed to account for the damping losses associated with the relative motion of the various components of the OoC. The damping chosen for the model is a fit to give reasonable BM passive gain relative to stapes. Dimensions and various angles used are estimated based on Refs. 24 and 28, the OoC sketch shown in Ref. 33, and data from Refs. 9 and 34.

We use a reference displacement at the stapes to solve for the acoustic response. Therefore, only the round window impedance is included in the model, for the rest of the middle ear is not required for the results presented in this study. The round window termination is modeled as a spring and damper interacting with the fluid in the lower duct. The fluid geometry is given in Table I and fluid density is taken as that of water (1000 kg/m^3). As a simplification, we use a uniform cross-section two-duct model with a cross-sectional area corresponding to a basal location in a guinea pig (the extension to a nonuniform area is straightforward, but we have chosen not to study this effect). The fluid model is

considered an approximation to the three-dimensional fluid effects. In order to model the true fluid geometry, one would have to build an intricate model of the sulcus and the sub-tectorial membrane spaces. Such a model would be computationally expensive. Our goal is to build the simplest model that represents the fundamental trends and mechanics of the passive and active response of the cochlea to acoustic stimulation.

B. Electrical parameters

The resistances and capacitances of the apical and basolateral portion of the OHC are known from measurements and model estimates.[35,36] The present model possesses the 1 kHz RC cut-off in OHCs for basal locations. The values shown for these quantities in Table III assume that there are 100 rows of OHCs per millimeter in a guinea pig.[9] The endocochlear potential is taken from Ref. 37. Various resistances in the model are chosen based on Ref. 37 although variation of the resistances along the length is not included. The longitudinal resistance in the scala tympani represents longitudinal resistance in the interstitial space in the OoC. Hence the value of resistance used for that cable is as per the value used in Ref. 37 for the OoC cable resistance. The electro-mechanical coupling coefficient (ϵ_3) along the cochlea is estimated from Ref. 38.

FIG. 5. BM magnitude at different locations in dB normalized to stapes motion (i.e., BM gain relative to stapes). Dash curves: passive ($G_a^1(0)$ $=0$ S/m^2) and solid curves: active ($G_a^1(0)=4.19\times10^6$ S/m^2). Locations are (from left to right): 11.75, 10.75, 9.75, 8.75, 7.75, 6.75, and 5.75 mm.
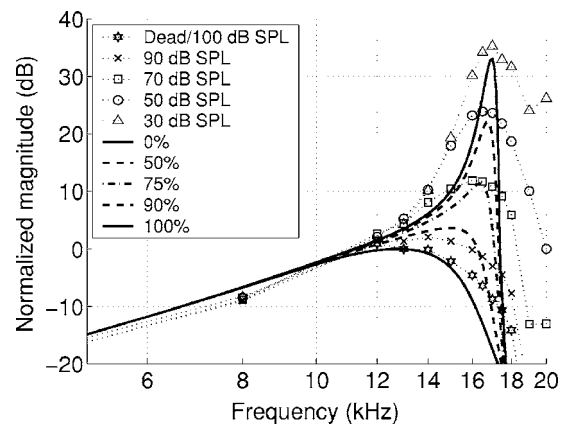
FIG. 7. Dotted lines with markers: Normalized guinea pig data from Cooper (Ref. 39). Thick lines: Theoretical prediction for BM magnitude in decibel at 6 mm for different levels of activity [100% activity is equivalent to a conductance slope of $G_a^1(0)=4.19\times10^6$ S/m^2]. Magnitudes are normalized with respect to maximum passive response.
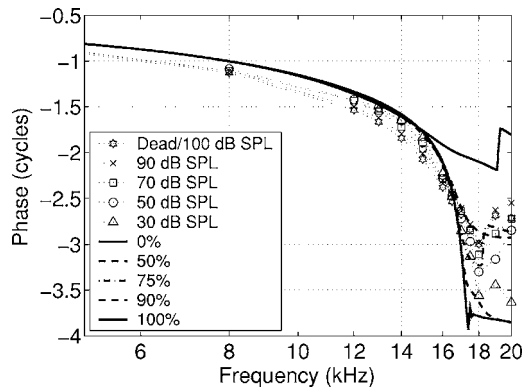
FIG. 8. Dotted lines with markers: Guinea pig data from Cooper (Ref. 39). Thick lines: BM phase in cycles (2π rad) at 6 mm for different levels of activity [100% activity is equivalent to a conductance slope of $G_a^1(0)$ =4.19×10^6 S/m^2].

FIG. 9. Dash lines: Normalized guinea pig data from deBoer (Ref. 40); higher gains correspond to lower dB SPLs. Solid lines: Theoretical prediction for BM magnitude in decibels at 6 mm for different levels of activity [100% activity is equivalent to a conductance slope of $G_a^1(0)$=4.19 $\times10^6$ S/m^2]. Higher gains correspond to higher conductance slope. Magnitudes are normalized with respect to maximum passive response.

The MET sensitivity (i.e., the relation between the HB rotation to the change in HB conductance) is kept as a free variable in the model. Having zero MET sensitivity is roughly equivalent to a passive cochlea. Increasing MET sensitivity increases the gain in the model. As expected, the model depends nonlinearly (and strongly) on the MET sensitivity. Indeed, if the MET sensitivity is increased beyond a critical limit, the model response becomes unstable. We found that different regions on the BM have different maximum values of the MET sensitivity such that the global response was stable. A polynomial-exponential dependence was used for the variation of the MET sensitivity in the model,

$$G_a^1(x) = G_a^1(0)(-10^5x^3 + 5.9\times10^3x^2 - 109x + 1.0)e^{-150x},$$
(36)

where x is in meters. Figure 4 shows the function $G_a^1(x)$ given in Eq. (36) plotted against x. This dependence was found from numerical experimentation and comparison of predictions with experiments. The maximum value for $G_a^1(0)$ at which the response of the model is stable is 4.19 $\times10^6$ S/m^2. Our empirically determined spatial variation for the MET channel sensitivity is used throughout the study. The entire spatial pattern for the MET sensitivity is scaled by the multiplying factor $G_a^1(0)$ to alter the activity level in the model. An analysis of whether this yields physiologically reasonable transducer currents is presented in Sec. IV D.

IV. RESULTS

The finite element model is solved for unit stapes displacement input which represents the acoustic input. A mesh discretization of 1041 nodes along x direction and 41 nodes along z direction was used for the model. The convergence and sensitivity to the discretization are discussed in Ref. 20. All results presented in this section are with respect to the stapes input and use three symmetric fluid modes and a feedforward distance of 5 μm. To alter the activity level we change the maximum MET channel sensitivity ($G_a^1(0)$) while leaving *all* other parameters constant. It is arguable whether reducing the MET sensitivity is equivalent to reduction in
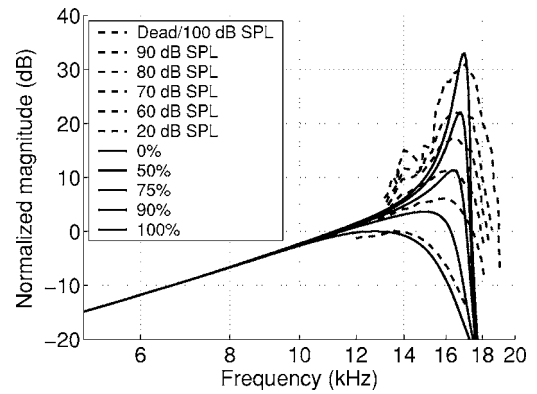
activity level. The other parameter controlling activity, ϵ_3, is not expected to change significantly over the small variation in transmembrane potential ($<$15 mV, see Ref. 2) seen *in vivo*. At higher acoustic levels the HB conductance would start saturating much earlier than ϵ_3 (if it does saturate at all). Therefore we approximate the higher sound input level by a lower MET sensitivity in the model. Naturally this is not strictly correct as we are ignoring nonlinearities in this model. In the results presented, $G_a^1(0)$=0 S/m^2 represents a passive case while $G_a^1(0)$=4.19×10^6 S/m^2 represents the fully active case [see Sec. III B and Eq. (36)]. The following shows the model's response predictions to acoustic input at varying locations, varying activity levels, response to impulse input, frequency glides exhibited by the model, and the BM motion relative to the ST potential (magnitude and phase). We start with the frequency response of the model at different locations along the BM.

A. Frequency response

1. Different locations along the BM

Figures 5 and 6 show magnitude and phase, respectively, of the BM frequency response referenced to stapes motion at different locations for high activity and no activity (passive model). The model exhibits familiar tuning curves from base to apex. The response shifts to lower frequencies by about half an octave when activity is lowered and amplitude drops by about 35 dB from active to passive for basal locations.

2. Different activity levels

Figures 7 and 8 show magnitude and phase, respectively, at x=6 mm for different activity levels overlapped with experimental results from Cooper [Fig. 1(a) in Ref. 39]. Figures 9 and 10 show the same model results overlapped with experimental results from deBoer and Nuttall.[40] In the magnitude plots, the BM gain with respect to stapes is normalized to the maximum passive/dead response. At frequencies lower than the best frequency, the model response (either phase or amplitude) is insensitive to alterations in the level of activity, consistent with experimental observations. The
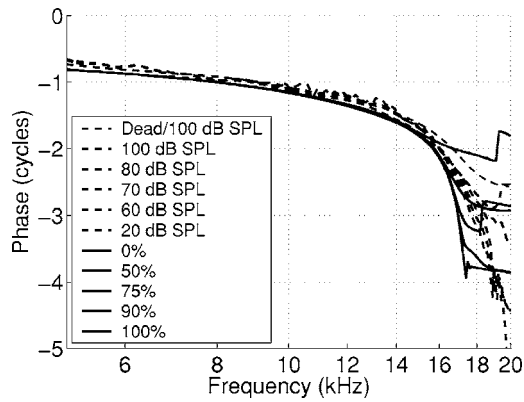
FIG. 10. Dash lines: Guinea pig data from deBoer (Ref. 40) (data supplied by E. deBoer). Thick lines: BM phase in cycles (2π rad) at 6 mm for different levels of activity [100% activity is equivalent to a conductance slope of $G_a^1(0)=4.19\times10^6$ S/m^2].

FIG. 11. Experimental results for the guinea pig BM gain envelope at the outer pillar (OP) location and the second (middle) row OHC location from Ref. 41. The thick lines represent the active response while thin lines represent the passive response. Notice that the maximum passive response depends on the radial location where it is measured.

close match between the model prediction and the experimental data for the frequency dependence of the BM gain is evident in Figs. 7 and 9. In comparing the two magnitude plots (Figs. 7 and 9), the variability in the experimental data should be noted. For example, the relative gain from 100 to 70 dB SPL is around 11 dB for data from Cooper[39] while the relative gain is around 17 dB for data from deBoer and Nuttall.[40] The model predicts a relative gain of approximately 35 dB from the passive state [represented by $G_a^1(0)$ =0 in the model] to the state with maximum activity. A similar relative gain is also seen in experiments as we transition from high intensity acoustic input (where the effect of active force generation in the cochlea is expected to be minor) to low intensity acoustic input (where the cochlea is at its most sensitive). The characteristic frequency (CF) of the location also shifts by around half an octave in both the model and the experiments with the reduction in activity, which is embodied by a lower MET sensitivity in the model and increased acoustic sound input level in the experiments. The phase comparisons in Figs. 8 and 10 too show a close match between the experiments and the model, although the passive response of the model does not accumulate as much phase lag as is seen in experiments at high sound levels.

Magnitude predictions of the model are sharper than the experimental values (i.e., Q_{10} dB for the model is higher than the experiment). The sharpness of response varies among experiments and does depend on processing of the data somewhat (as is evident from Figs. 7 and 9). Even including such experiment-to-experiment variability, the model exhibits a higher Q_{10} factor than seen in a typical experiment. This sharp response also results in a longer impulse response as shown in Sec. III B. We speculate that the main reason for this discrepancy is the lack of longitudinal structural coupling along the BM and the TM in the model. Although we have not directly compared absolute gains, our model predicts the passive response [$G_a^1(0)=0$ S/m^2] to be around 10 dB higher than both the results in Refs. 39 and 40. The passive gain of the model is around 33 dB (see Fig. 5). It should be noted that the magnitude observed in experiments depends on the location of the bead on the BM.[22] Passive gain as high as 40 dB has been measured in experiments
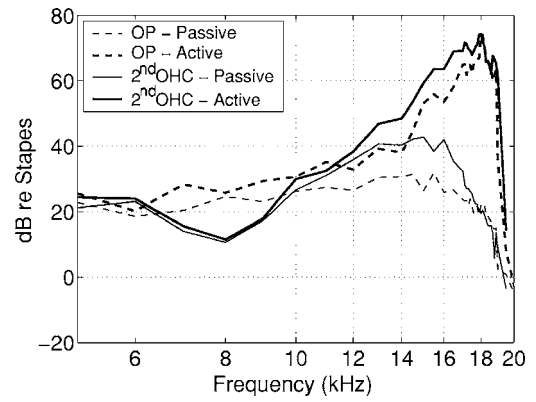
(results reproduced in Fig. 11). Envelopes of guinea-pig BM response data from Ref. 41 for acoustic levels varying from 8 to 100 dB SPL at two different radial locations are shown in Fig. 11 to demonstrate the spatial variability of the BM response. As Fig. 11 shows, the measured passive gain of the BM is as high as 40 dB when measured near the middle OHC location but is only 30 dB when measured at an outer pillar location.

B. Impulse response

The impulse response of the model was obtained by first simulating the model at different frequencies (100 Hz to 20 kHz in 100 Hz steps) and then applying the inverse Fourier transform at a given location. Figure 12 shows the impulse response of the model for an active and a passive case
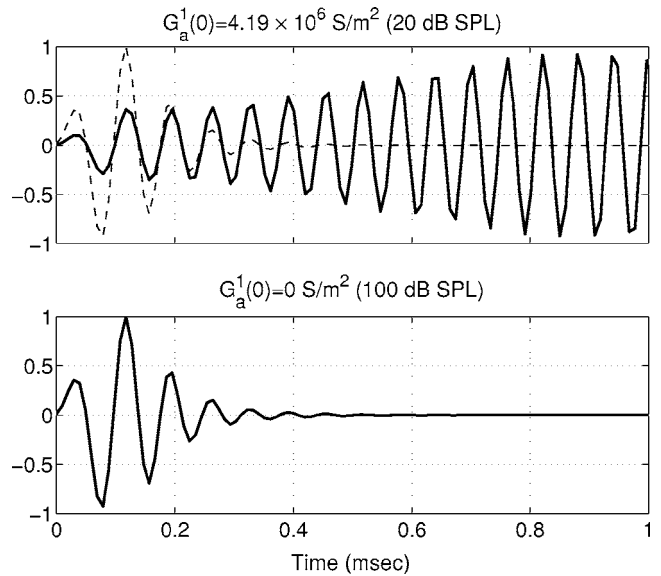
FIG. 12. Normalized model predictions of the BM impulse response at 6 mm for a highly active and a passive case. Active case corresponds roughly to gain seen at 20 dB SPL, while passive case corresponds to gain seen at 100 dB SPL or in a dead cochlea. The thin dashed line in the top panel is the 100 dB SPL result superimposed to highlight the near invariance of the zero crossings of the response.

Ramamoorthy et al.: Modeling acoustic response of the cochlea
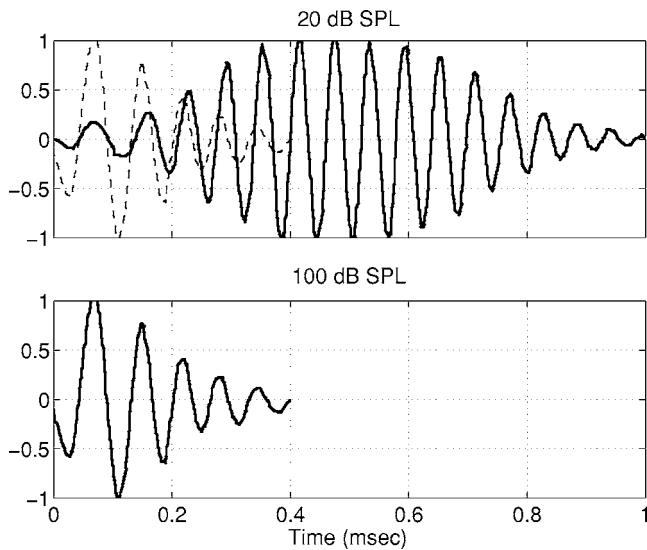
FIG. 13. Normalized experimental BM impulse response as shown in Ref. 40. Acoustic levels for the two plots are 20 and 100 dB SPL (dead cochlea). The thin dashed line in the top panel is the 100 dB SPL result superimposed to highlight the near invariance of the zero crossings of the response.

FIG. 15. BM response magnitude at 6 mm in decibels with respect to stapes motion for active [thick lines, $G_a^1(0)=4.19\times10^6$ S/m^2] and passive case [thin line, $G_a^1(0)=0$ S/m^2], with and without longitudinal cables. The passive curves virtually overlap and so only one of them is plotted.

(roughly corresponding to a 20 and a 100 dB SPL case as per Fig. 9) and Fig. 13 shows the experimentally derived impulse response as seen by Ref. 40 in a guinea pig. The passive response of the model compares well with the experiment but the active response shows oscillations that continue on up to 3 ms (the figure shows response only up to 1 ms to aid comparison between model and experiment). This would be expected as the model has a sharper frequency response than the experimental data. As in the experimental data (e.g., Ref. 40 and in the theoretical study by Ref. 42), the model also shows the same zero crossings of the response for the initial few cycles (indicated by the passive response superimposed as thin dashed line over the active response). The model does not exhibit any ringing behavior often seen in experimental data[43] and also reproduced by nonlinear models.[26] This suggests that ringing could be primarily due to nonlinear effects in the cochlea.

Figure 14 shows the frequency glide for the active impulse response shown in Fig. 12. The glide represents the

FIG. 14. Frequency glide at 6 mm for the most active case ($G_a^1(0)=4.19 \times 10^6$ S/m^2) which roughly corresponds to a 20 dB SPL acoustic input as per Fig. 9 for the 17 kHz CF location.

instantaneous frequency at a given time of the impulse response. The instantaneous frequency is much lower than the CF of the location at the start of the impulse response but grows rapidly to reach the CF. This result is very similar to the glides observed experimentally (e.g., Refs. 43 and 44).

C. Removal of cables

One interesting test that can be done using the model is to evaluate the importance of current conduction along the length of the cochlea to the BM amplification. To run this test we disabled the longitudinal cables and solved for the system response to acoustic input.

Figure 15 shows the absence of the cables has a measurable impact on cochlear amplification when activity is high. The effect is negligible when the model is passive. There is a loss in amplification of around 10 to 15 dB when cables are removed. The model results imply that current conduction along the length of the cochlea also plays a role in active amplification in the cochlea.

D. Potentials and currents

We have shown so far that the mechanical response of the BM compares favorably with experimental results. The validity of the model results depends on whether the voltages in the different scalae and the currents predicted by the model are within certain reasonable limits. Voltages or currents too large or too small compared to experimentally measured quantities would indicate possible shortcomings of the model. This model has been constructed with an aim to test efficacy of OHC forcing at high frequencies. The model produces high amplification at the base like in experiments using only active somatic forcing from OHCs. We therefore report the model predictions at a basal location and compare those currents and voltages with experimental values. Since the model is linear, the model predictions should be compared with very low level acoustic data. We use a BM displacement of 0.5 nm as a reference for stating the voltages

and currents assuming that the cochlear response shows a linear input-output relation up to around 0.5 nm BM displacement (see Fig. 3 in Ref. 45). At the 17 kHz best place, data from Ref. 41 suggest that a 10 dB SPL acoustic input will give a 0.5 nm BM response while that from Ref. 39 suggest that a 30 dB SPL acoustic input is needed to achieve the same. A 0.5 nm displacement of the BM is therefore considered to roughly correspond to a 20 dB SPL sound input.

With the highest value of MET sensitivity [$G_a^1(0) = 4.19 \times 10^6$ S/m^2, or at $x=6$ mm, $G_a^1(0.6)=0.91 \times 10^6$ S/m^2] the model predicts a transducer current of 1.56 nA and a transmembrane potential of 0.62 mV for 0.5 nm BM displacement at the 17 kHz best place ($x=6$ mm). Corresponding predicted maximum displacements for HBs and OHCs at $x=6$ mm are 2.5 and 3.5 nm respectively, per nm of BM displacement. The maximum predicted active force from one OHC on the BM at $x=6$ mm is 0.12 nN/nm of BM displacement. Recent estimates for transducer currents in gerbils are available in Ref. 12. The maximum current measured by He et al.[12] at a basal location in a hemicochlea preparation was around 2.5 nA. According to corrections presented in Ref. 13 for reduced endocochlear potential, reduced body temperature of the animal, and differences in endolymphatic cations, the current in vivo is expected to be around four times higher than what was observed in Ref. 12. This gives us an estimate of 10 nA for the maximum transducer current. A review by Robles et al.[45] indicates that typically transduction channels are saturated at 100 dB SPL sound input (as inferred from the linear growth in BM response beyond 100 dB SPL). Guinea pig data from Refs. 41 and 39 suggest that at 100 dB SPL acoustic input, the BM displacement is less than 20 nm. Making a linear approximation of the transducer current to BM displacement relation yields a sensitivity of 0.5 nA/nm of BM displacement. This is an extremely conservative estimate since the MET sensitivity is expected to be highest (or close to its highest value) near the resting state and greatly reduced as the channels saturate.[9] The transducer current sensitivity therefore is likely to be much higher than 0.5 nA/nm. This model for the most sensitive case at the 6 mm site (the 17 kHz best place), predicts a transducer current of 3.12 nA/nm of BM displacement, which is around six times higher than the 0.5 nA/nm lower bound estimated earlier. To summarize, for the most sensitive model, our predictions of HB currents (1.56 nA at ≈20 dB SPL) are lower than the maximum values given in the literature.

The ST voltage

In an interesting experiment, Fridberger et al.[46] measured OoC potentials in response to acoustic stimulation. They observed that ST voltage and BM velocity profiles are very similar in acoustic stimulation and the phase difference between ST and BM has a sharp increase near the best place. We compare our model predictions of the ST voltage with the experimental data of Fridberger et al. (2004). The location chosen is $x=5.85$ mm, which has a CF of 17.4 kHz, the same as the CF of the location in the experiment of Ref. 46.
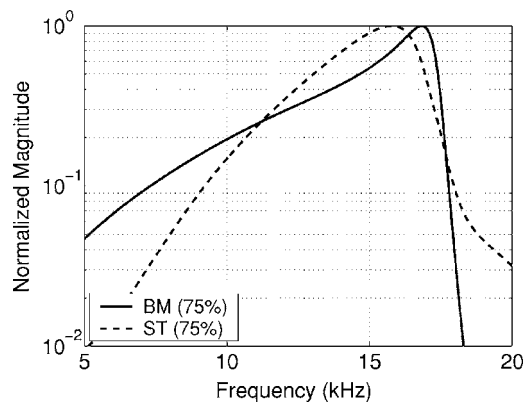
FIG. 16. Model prediction of the magnitudes of the BM displacement and ST voltage at 75% activity [$G_a^1(0)=3.14 \times 10^6$ S/m^2] normalized to their respective maximum levels, plotted vs frequency of the acoustic stimulus. Location is $x=5.85$ mm (CF ≈ 17.4 kHz).

Experimental results from Ref. 46 show that the BM displacement and OoC voltage have nearly the same frequency dependence when normalized to their maximum response at the CF (experimental results for magnitude not reproduced here). For the most active case [i.e., when the MET sensitivity is maximum $G_a^1(0)=4.19 \times 10^6$ S/m^2], our model predicts that the BM displacement is much more sharply tuned than the ST voltage. When a lower level of activity (as embodied by a lower MET sensitivity) is used the normalized profiles of ST voltage and BM displacement match each other more closely as in the experiments (see model prediction for 75% activity shown in Fig. 16). In terms of absolute magnitude, the model predicts 0.17 mV for 1 nm of BM displacement (0.24 mV rms) for the lower MET sensitivity [$G_a^1(0)=3.14 \times 10^6$ S/m^2, or 75% activity], while Ref. 46 shows around 0.15 mV rms ST potential when BM displacement is 1 nm at 40 dB SPL. Model results for the phase difference between BM displacement and ST voltage (Fig. 17) are in good agreement with experimental measurements from Ref. 46. The phase difference is nearly constant until near the CF where it jumps over 300° for the most
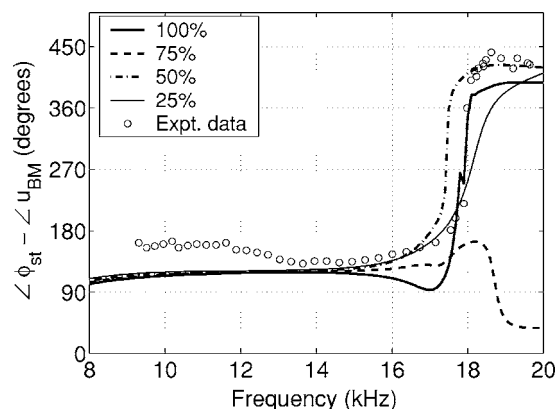
FIG. 17. Phase difference between BM displacement and ST voltage plotted vs frequency of acoustic stimulus for different activity levels [100% activity is equivalent to a conductance slope of $G_a^1(0)=4.19 \times 10^6$ S/m^2]. Location is $x=5.85$ mm (CF ≈ 17.4 kHz both in experiment and model). Experimental data from Ref. 46.

Ramamoorthy et al.: Modeling acoustic response of the cochlea
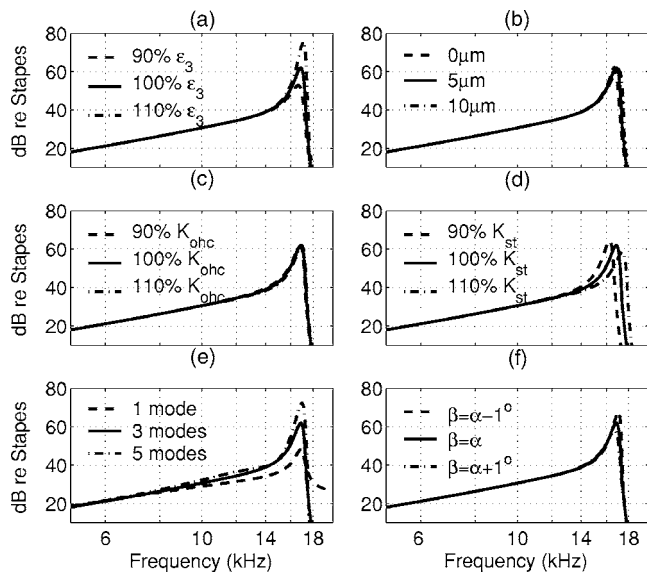
FIG. 18. Gain of the BM displacement compared for different: (a) OHC electromechanical coupling coefficient (ϵ_3), (b) feed-forward distance, (c) OHC stiffness (K_{ohc}), (d) HB stiffness (K_{hb}), (e) number of fluid modes, and (f) HB angle (angle β, see Fig. 3). Location is $x=6$ mm (CF $\approx$ 17 kHz), $G_a^1(0)=3.98 \times 10^6$ S/m^2 or 95% active.

active case. As was observed by Ref. 46, the jump does not increase in a monotonic manner with increasing acoustic level.

The results from Ref. 46 were suggested to be in support of a theory on OHC somatic motility proposed by Dallos and Evans.[36] The Dallos–Evans theory is that OHC somatic motility is driven not by fluctuations in the transmembrane voltage, but by fluctuations in the extracellular voltage. Our model has OHC somatic motility driven by fluctuations in the transmembrane voltage and yet the model predicts results similar to those seen in Ref. 46. The use of data from Ref. 46 in favor of the Dallos–Evans theory is therefore questionable.

V. PARAMETER SENSITIVITY

In the following, we investigate the effect of varying a few of the key parameters in the model that are expected to influence the BM motion significantly. The sensitivity of the model to the MET conductance is evident from the results presented in the preceding results section. In Fig. 18 we show the change in the model response when small changes are made in HB stiffness (K_{hb}), OHC stiffness (K_{ohc}), OHC electromechanical coupling coefficient (ϵ_3), HB orientation in the cochlea (angle β), number of fluid modes, and the feed-forward distance. The conductance slope was set at 95% of the maximum value for all plots in Fig. 18 ($G_a^1(0)=3.98 \times 10^6$ S/m^2) to avoid the model transitioning into an unstable region.

A. Electromechanical coupling

The model exhibits higher cochlear amplification for a higher electromechanical coupling coefficient [Fig. 18(a)]. The model response is sensitive to this parameter with over 5 dB change in BM gain when the electromechanical cou-

pling parameter is changed by 10%. A higher electromechanical coupling coefficient implies higher force generation by OHCs so this result is expected.

B. Feed-forward distance

Unlike the model developed by Lim and Steele,[26] this model is not highly sensitive to the forward inclination angle of the OHCs [Fig. 18(b)]. Lim and Steele[26] proposed that the forward inclination of OHCs plays a significant role in cochlear tuning and amplification. They use a gain factor to account for the feed-forward effect. In this model, the effect of the forward inclination of OHCs is negligible. Nonetheless we cannot rule out the possibility that if the pressure forcing on the TM is accurately modeled, it might show that effect of feed-forward is larger than what is seen here. Furthermore, a different selection of parameters (values and functional dependence) and the inclusion of longitudinal structural coupling may influence this conclusion.

C. OHC stiffness

Changing OHC stiffness by 10% has a very small effect on the BM response [Fig. 18(c)]. Increased OHC stiffness gives a higher gain in the BM response with respect to the stapes while a lower OHC stiffness yields a smaller BM response (not discernible in the figure). This is consistent with the implication in Ref. 4 that increased OHC stiffness leads to higher cochlear amplification. This result should be interpreted with caution though. It is likely that other factors (such as the OHC basolateral capacitance and electromechanical coupling coefficient) were also affected in Ref. 4 along with the OHC stiffness when the current was injected in the cochlea, so a direct comparison is not possible unless all affected parameters in the experiment can be identified.

D. HB stiffness

A change in the HB stiffness causes a shift in the characteristic frequency of the location [Fig. 18(d)]. In the model the HB stiffness dominates stiffness loading on the TM due to which the resonance mode of the OoC shifts when this stiffness is altered. This points to a more prominent role for HBs if HB motility is present.

E. Fluid modes

For the box model of the cochlea, increasing the number of fluid modes represents the transition from a two-dimensional cochlear model (one mode) to a three-dimensional model (more than one mode). However, since the shape of the cross section in the cochlea is unlike a rectangular box, the results shown here should only be taken as indicative of the behavior possible in a cochlear model with a more realistic geometry. Figure 18(e) shows the model prediction for magnitude of the BM response with 95% activity [i.e., $G_a^1(0)=3.98 \times 10^6$ S/m^2] at $x=6$ mm with five, three, and one fluid mode. As the results show, having more fluid modes influences the rate of amplitude drop after the best place and also improves active amplification in the model. The passive cut-off behavior also improves (not shown)

when higher modes are used and the passive curves shift to more lower frequencies. However, the passive gain is largely unaffected by the number of fluid modes.

F. HB angle

To study the role played by the kinematic configuration used in the model one can simulate the model with small perturbations to the different angles and lengths of the structures. The most critical kinematic feature of them is the relative angle between the HBs and the OHCs. Figure 18(f) shows BM response when HB angles (β) are changed slightly keeping the OHC angles (α) the same. A very small change was used since the model is at high activity $[G_a^1(0) = 3.98 \times 10^6 \text{ S/m}^2]$ and can be easily made unstable if large changes in angle are used. If the HB angle is smaller than OHC angle, the model predicts a higher BM gain. If one looks at the deformation statically, the upward motion of the BM would result in larger rotation of the HBs in the excitatory direction and therefore would result in more transducer current and more OHC forcing if the HB angle is smaller than OHC angle. Having all the HBs "leaned back" in the direction of the tallest HB would thus give better BM amplification.

VI. CONCLUSION

The main purpose of the present paper was to describe and vet a mathematical model of the cochlea. In this paper, we have presented a physiologically motivated mechanical-electrical-fluidic model for predicting the response of the cochlea to acoustical stimulus. The model explicitly couples the mechanical and electrical degrees of freedom through a piezoelectric model of OHC electromotility and includes the conduction of electrical energy down the length of the cochlea. HB motility is not included in this model. We have shown that by changing a single parameter, the MET sensitivity, the model captures a wide range of experimentally observed effects in the cochlea. The model predictions of the transducer currents, receptor potentials, and mechanical response (BM displacement) compare closely to available experimental data. Under a variation of the model parameters, we have shown that the qualitative nature of the predictions do not change dramatically, although the details of the response certainly do. We do not claim that the model gives a complete picture of cochlea. Key simplifications include the linearization of all response variables about their operating points, and the simplification of the fluid geometry. We do claim to have made progress on describing how electromotile processes might work in the cochlea, especially in the basal portion. The model brings to light the following two important points in cochlear mechanics:

1. Modes of vibration of the OoC: What is thought of as the resonance of the BM should be interpreted as the local resonance of the OoC system. Our simplified kinematical approximation to the OoC system suggests, as does experimental work,[32] that structures of the OoC other than the BM, especially the TM, play a prominent role in the cochlea. Observations of a notch in the BM frequency

response to bipolar stimulation[8] are indicative of a resonance in the OoC and give further credence to this assertion.

2. Intimate link between the electrical currents/voltages and OoC micromechanics: The elements in the OoC together with the electrical environment form a complex feedback system. To accurately model the feedback force arising from the OHCs or HBs it is essential to have all the individual elements modeled in the system. In addition to providing for a predictive model, the availability of current and voltage responses provide additional data to verify the validity of the model.

It is emphasized here that the highest RC-cutoff frequency is around a 1000 Hz for the basal cells in the model. The model predictions for the forces generated by OHCs, the receptor potential, the OoC potential (equivalent to the ST potential in our model), and the transducer currents all fall below maximum measured values. Hence the model developed herein shows how an electromechanical system can interact to overcome filtering of the transmembrane potential by the membrane capacitance. The MET sensitivity (probably most reliably quantified experimentally *in situ* by the BM displacement to transducer current measurement as HB motion is difficult to measure in that setting) used in the model is likely higher than the sensitivity *in vivo*. As the results in Sec. V indicate, the model response is sensitive to parameters and a perhaps further optimization of the parameters could lead to a lower estimate of the maximum MET sensitivity. We conjecture here that it is quite possible that a higher MET sensitivity is needed in the present model to overcome the lack of HB motility in the model. The HBs are of course ideally situated to aide in channel opening and thereby affecting the conductance of the channels. Further experimental and theoretical work is needed to validate such a hypothesis. It is therefore claimed that HB forcing, if present, acts in conjunction with the OHC somatic forcing to give rise to cochlear amplification as suggested in a recent finding by Kennedy *et al.*[47]

[1] G. V. Bekesy, *Experiments in Hearing* (McGraw Hill, New York, 1960).
[2] P. Dallos, J. Santos-Sacchi, and A. Flock, "Intracellular recordings from cochlear outer hair-cells," Science **218**, 582–584 (1982).
[3] A. Hubbard and D. C. Mountain, "Alternating current delivered into the scala media alters sound pressure at the eardrum," Science **222**, 510–512 (1983).
[4] A. A. Parthasarathi, K. Grosh, J. F. Zheng, and A. L. Nuttall, "Effect of current stimulus on in vivo cochlear mechanics," J. Acoust. Soc. Am. **113**, 442–452 (2003).
[5] W. E. Brownell, C. R. Bader, D. Bertrand, and Y. D. Ribaupierre, "Evoked mechanical responses of isolated cochlear hair cells," Science **227**, 194–196 (1985).
[6] J. F. Zheng, T. Y. Ren, A. Parthasarathi, and A. L. Nuttall, "Quinine-

2772   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Ramamoorthy *et al.*: Modeling acoustic response of the cochlea

induced alterations of electrically evoked otoacoustic emissions and cochlear potentials in guinea pigs," Hear. Res. **154**, 124–134 (2001).

[7] G. Frank, W. Hemmert, and A. W. Gummer, "Limiting dynamics of high-frequency electromechanical transduction of outer hair cell," Proc. Natl. Acad. Sci. U.S.A. **96**, 4420–4425 (1999).

[8] K. Grosh, J. F. Zheng, Y. Zou, E. de Boer, and A. L. Nuttall, "High-frequency electromotile responses in the cochlea," J. Acoust. Soc. Am. **115**, 2178–2184 (2004).

[9] P. Dallos, "Overview: Cochlear neurobiology," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York, 1996).

[10] P. Martin, D. Bozovic, Y. Choe, and A. J. Hudspeth, "Spontaneous oscillation by hair bundles of the bullfrog's sacculus," J. Neurosci. **23**, 4533–4548 (2003).

[11] H. J. Kennedy, A. C. Crawford, and R. Fettiplace, "Force generation by mammalian hair bundles supports a role in cochlear amplification," Nature (London) **433**, 880–883 (2005).

[12] 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).

[13] H. J. Kennedy, M. G. Evans, A. C. Crawford, and R. Fettiplace, "Fast adaptation of mechanoelectrical transducer channels in mammalian cochlear hair cells," Nat. Neurosci. **6**, 832–836 (2003).

[14] S. Neely, "A model of cochlear mechanics with outer hair cells motility," J. Acoust. Soc. Am. **94**, 137–146 (1993).

[15] C. D. Geisler and C. Sang, "A cochlear model using feed-forward outer-hair cells forces," Hear. Res. **86**, 132–146 (1995).

[16] P. J. Kolston, "Comparing in vitro, in situ, and in vivo experimental data in a three-dimensional model of mammalian cochlear mechanics," Proc. Natl. Acad. Sci. U.S.A. **96**, 3676–3681 (1999).

[17] K. M. Lim and C. R. Steele, "A three-dimensional nonlinear active cochlear model analyzed by the wkb-numeric method," Hear. Res. **170**, 190–205 (2002).

[18] 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).

[19] E. K. Dimitriadis and R. S. Chadwick, "Solution of the inverse problem for a linear cochlear model: A tonotopic cochlear amplifier," J. Acoust. Soc. Am. **106**, 1880–1892 (1999).

[20] 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).

[21] R. C. Naidu and D. C. Mountain, "Measurements of the stiffness map challenge a basic tenet of cochlear theories," Hear. Res. **124**, 124–131 (1998).

[22] M. Homer, A. Champneys, G. Hunt, and N. P. Cooper, "Mathematical modeling of the radial profile of basilar membrane vibrations in the inner ear," J. Acoust. Soc. Am. **116**, 1025–1034 (2004).

[23] W. Rall, "Core conductor theory and cable properties of neurons," in *Handbook of Physiology: The Nervous System*, edited by E. R. Kandel, J. M. Brookhardt, and V. B. Mountcastle (Williams and Wilkins, 1997), Vol. **1**, Chap. 3, pp. 39–98.

[24] P. Dallos, "Organ of corti kinematics," J. Assoc. Res. Otolaryngol. **4**, 416–421 (2003).

[25] N. V. Deo and K. Grosh, "Simplified nonlinear outer hair cell models," J. Acoust. Soc. Am. **117**, 2141–2146 (2005).

[26] K. M. Lim and C. R. Steele, "Response suppression and transient behavior in a nonlinear active cochlear model with feed-forward," Int. J. Solids Struct. **40**, 5097–5107 (2003).

[27] 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).

[28] C. Fernandez, "Dimensions of the cochlea (guinea pig)," J. Acoust. Soc. Am. **24**, 519–523 (1952).

[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] D. Z. Z. He and P. Dallos, "Properties of voltage-dependent somatic stiffness of cochlear outer hair cells," J. Assoc. Res. Otolaryngol. **1**, 64–81 (2000).

[31] J. J. Zwislocki and L. K. Cefaratti, "Tectorial membrane. II Stiffness measurement in vivo," Hear. Res. **42**, 211–228 (1989).

[32] A. W. Gummer, W. Hemmert, and H. P. Zenner, "Resonant tectorial membrane motion in the inner ear: Its crucial role in frequency tunings," Proc. Natl. Acad. Sci. U.S.A. **93**, 8727–8732 (1996).

[33] K. E. Nilsen and I. J. Russell, "The spatial and temporal representation of a tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **97**, 11751–11758 (2000).

[34] A. A. Spector, W. E. Brownell, and A. S. Popel, "Effect of outer hair cell piezoelectricity on high-frequency receptor potentials," J. Acoust. Soc. Am. **113**, 453–461 (2003).

[35] G. D. Housley and J. F. Ashmore, "Ionic currents of outer hair-cells isolated from the guinea-pig cochlea," J. Physiol. (London) **448**, 73–98 (1992).

[36] P. Dallos and B. N. Evans, "High-frequency motility of outer hair cells and the cochlear amplifier," Science **267**, 2006–2009 (1995).

[37] D. Strelioff, "Computer-simulation of generation and distribution of cochlear potentials," J. Acoust. Soc. Am. **54**, 620–629 (1973).

[38] K. H. Iwasa and M. Adachi, "Force generation in the outer hair cell of the cochlea," Biophys. J. **73**, 546–555 (1997).

[39] 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).

[40] 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).

[41] 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).

[42] C. A. Shera, "Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics," J. Acoust. Soc. Am. **110**, 332–348 (2001).

[43] A. Recio, N. C. Rich, S. S. Narayan, and M. Ruggero, "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989 (1998).

[44] E. de Boer and A. L. Nuttall, "The mechanical waveform of the basilar membrane. I. Frequency modulations (glides) in impulse responses and cross-correlation functions," J. Acoust. Soc. Am. **101**, 3583–3592 (1997).

[45] L. Robles and M. A. Ruggero, "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352 (2001).

[46] A. Fridberger, J. B. de Monvel, J. Zheng, N. Hu, Y. Zou, T. Ren, and A. Nuttall, "Organ of corti potentials and the motion of the basilar membrane," J. Neurosci. **24**, 10057–10063 (2004).

[47] H. J. Kennedy, M. G. Evans, A. C. Crawford, and R. Fettiplace, "Depolarization of cochlear outer hair cells evokes active hair bundle motion by two mechanisms," J. Neurosci. **26**, 2757–2766 (2006).

Scala vestibuli pressure and three-dimensional stapes velocity measured in direct succession in gerbil[a]

W. F. Decraemer[b]
University of Antwerp, CGB, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

O. de La Rochefoucauld, W. Dong, and S. M. Khanna
Columbia University, 630 West 168th Street, New York, New York 10032

J. J. J. Dirckx
University of Antwerp, CGB, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

E. S. Olson
Columbia University, 630 West 168th Street, New York, New York 10032

It was shown that the mode of vibration of the stapes has a predominant piston component but rotations producing tilt of the footplate are also present. Tilt and piston components vary with frequency. Separately it was shown that the pressure gain between ear canal and scala vestibuli was a remarkably flat and smooth function of frequency. Is tilt functional contributing to the pressure in the scala vestibuli and helping in smoothing the pressure gain? In experiments on gerbil the pressure in the scala vestibuli directly behind the footplate was measured while recording simultaneously the pressure produced by the sound source in the ear canal. Successively the three-dimensional motion of the stapes was measured in the same animal. Combining the vibration measurements with an anatomical shape measurement from a micro-CT (CT: computed tomography) scan the piston-like motion and the tilt of the footplate was calculated and correlated to the corresponding scala vestibuli pressure curves. No evidence was found for the hypothesis that dips in the piston velocity are filled by peaks in tilt in a systematic way to produce a smooth middle ear pressure gain function. The present data allowed calculations of the individual cochlear input impedances. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2709843]

I. INTRODUCTION

The motion of the stapes footplate defines the mechanical middle ear output and produces the acoustical input to the cochlea. The mode of stapes vibration (piston-like and/or rocking) has been studied mainly in human temporal bones [for an overview see Heiland *et al.* (1999) and Hato *et al.* (2003)] and in a few animal models [anesthetized cat, Guinan and Peake (1967); cat temporal bone, Decraemer *et al.* (2003) and Decraemer and Khanna (2003); gerbil, Olson and Cooper (2000)]. Older measurements made use of capacitive probes, stroboscopic microscopy, and time-averaged holography, while the more recent measurements all made use of heterodyne laser interferometers, improving measurement quality and expanding the frequency range of measurements to most of the audible range. It is now well established that the motion is not purely piston-like but that at higher frequencies rotations about the long and short foot-plate axis are also present ("rocking," "tilting"). Whether the nonpiston motion components are functionally important is an open question. An argument in favor is the remarkable ability of the middle-ear to produce a gain between the sound pressure in the ear canal and the acoustic pressure in the perilymph of the scala vestibuli (directly behind the footplate) that is a very smooth function of frequency (Olson, 1998; Puria *et al.*, 1997) while the frequency response of the stapes exhibits pronounced maxima and minima. Could it be that tilting of the footplate of the stapes fills in the gaps in the frequency response of the piston component resulting in a smoother pressure frequency response in the cochlea? (Let us at this point make clear that when we call a function "smooth" in this paper we mean "with small differences between contiguous data points"). In an attempt to answer this question we have measured the scala vestibuli (SV) pressure using a micropressure sensor (Olson, 1998, 2001) followed by a measurement of the three-dimensional (3D) vibration velocity of the stapes in the adult gerbil using a confocal heterodyne microscope/interferometer (Decraemer and Khanna, 1999). We show the results for three animals. The pressure results are presented quite directly. The motion results require analysis to present them in the most useful coordinate system, and the issue of propagation of experimen-

tal error is presented in some detail. Finally we combine the pressure and velocity measurements, to find the cochlear input impedance.

II. MATERIAL AND METHODS

A. Animal preparation

Measurements were performed in three gerbils, 50–70 g in mass. The animal was first sedated with ketamine (1 mg/kg) and then deeply anaesthetized with sodium pentobarbital (initial dose 60 mg/kg). Supplemental doses (10 mg/kg) were given when a toe pinch response was elicited (checked at least every 20 min). The body temperature was maintained at 37 °C using an animal blanket. The animal head was firmly attached to a head holder using dental cement and surgery was performed to expose the left cochlea. A tracheotomy was performed to maintain a clear airway. Due to the length of the two-stage experiment (first SV pressure, then stapes motion) and the necessity of manipulating the animal when bringing it from one setup to the other, the animals did not survive during the entire second part of the experiment. The experiment was, without interruption, continued on the fresh cadaver. Animal 20m4 died during the last few runs of the 3D experiment (about 10 min before completion of the ~2.5 h experiment), animal 25m5 died after about 1 h and we restarted the entire 3D measurement. Animal 27m4 died after preparative surgery, all measurements were done on the fresh cadaver. The just post-mortem condition did not produce SV pressure or stapes velocity results that were significantly different.

The care and use of animals were approved by the Institutional Animal Care and Use Committee of Columbia University.

B. Sound system and pressure measurement in the ear canal

The external ear canal (EC) was cut short to its bony end and a small plastic tube with a diameter comparable to that of the EC was cemented in place. A sound delivery tube coupled to a sound source was tightly inserted in the plastic insert. A probe tube microphone was inserted coaxially within this tube and protruded slightly from the end within the EC. It measured the sound about 1 mm lateral from the superior edge of the pars tensa. Details of the sound source, sound delivery tube with coaxial probe microphone and microphone calibration have been described in detail previously (Khanna and Stinson, 1985).

The same sound driver and probe microphone were first used during the SV pressure measurements and later for the stapes velocity measurements. They had to be removed and replaced between the two stages of the experiment. Input to the sound driver was kept constant which resulted in a sound level in the EC that fluctuated slightly around 90 dB SPL.

C. Pressure measurements in the scala vestibuli

The intracochlear pressure in the SV was measured using a fiber-optic based micropressure sensor (diameter 170 μm) whose construction and calibration were described in Olson (1998). The sensitivity of the sensors is flat to within a few decibels up to at least 40 kHz. The sensors are calibrated the day of the experiment but the sensitivity is somewhat uncertain, as changes in sensor sensitivity from day to day or after an experiment of ~10 dB are not uncommon. These changes are flat with frequency, but set a limit on the accuracy of the absolute pressure. The sensitivity changes are likely a result of the fragility of the sensor membrane. To measure the pressure in the SV a small hole, the size of the pressure sensor, was hand drilled in the cochlear wall (Olson, 1998, Fig. 3). The pressure sensor was inserted to a depth of 100–200 μm with the aid of a micromanipulator. The distance with respect to the footplate was about 0.3 mm. This estimate was made later on a 3D model of the cochlea with inserted sensor, based on a micro CT scan of one of the temporal bones harvested at the end of the experiment (Dong and Olson, 2006, Fig. 1).

D. Hearing threshold measurements

An electrode was positioned at the round window opening to measure the compound action potential (CAP) response of the auditory nerve to tone pips as a measure of cochlear condition (Johnstone et al., 1979). The CAP response was measured prior to other measurements in the live animals 20m4 and 25m4. Both showed a starting CAP threshold curve that was typical for normal hearing animals with thresholds of 20–30 dB SPL over the most sensitive hearing range (0.5–22 kHz). Above 22 kHz the thresholds of the two animals diverged, and variability is typical of this vulnerable region of the cochlea. In animal 20m4, the 35 kHz threshold was 60 dB, in 25m4 the 40 kHz threshold was 40 dB.

E. Simultaneous stapes velocity and ear canal pressure measurements

After completion of the SV pressure measurement the sensor was removed. The tiny hole in the cochlea was left open. The hole in the middle ear cavity used for hearing threshold measurement was made wide enough (about 5–6 mm in a direction parallel to the annulus) to allow a good view of the stapes. As we see in Fig. 1 the stapes head and the 2/3rd upper part of the crura are visible. Note that with this approach we observed the stapes in the living gerbil from almost the opposite direction to that used for previous stapes measurements in cat temporal bone (Decraemer and Khanna, 2000; Decraemer et al., 2000). The approach allowed for a variation in observation angle of about ±15°, which is sufficient to determine the 3D components of the stapes motion. To prevent drying out of the specimen while the cavity was widely opened, small pieces (about 2 by 3 mm^2) of paper towel soaked in water were placed against the inner middle ear cavity wall in the space under the malleus handle. We took great care not to touch vibrating structures. In a later experiment during the motion measurement the specimen was also kept in the mist of an ultrasonic humidifier conducted via a plastic hose (diameter 1 cm) to
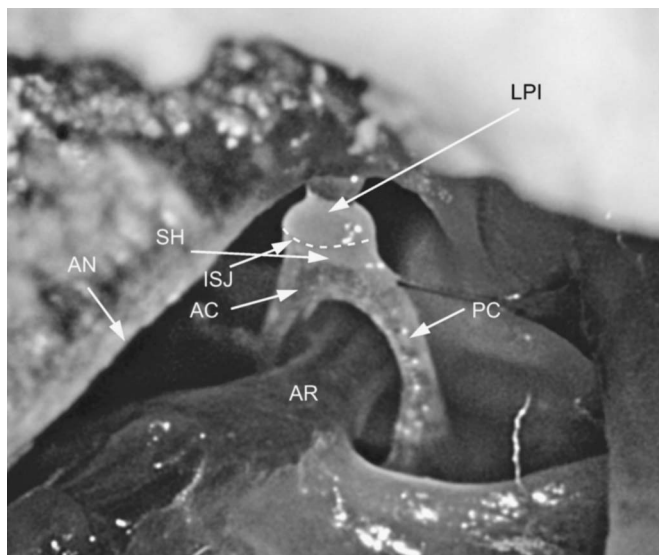
FIG. 1. A view of the stapes through a hole in the posterior ventral part of the bulla wall: LPI: lenticular process of the incus; SH: stapes head; ISJ: position of the incudo-stapedial joint; AC: anterior crus; PC: posterior crus; AN: part of the tympanic annulus; AR: artery running between the crura and blocking the view of the footplate.

about 2 cm from the opening in the bulla wall. These extra precautions did not markedly increase the already good stability of the preparation.

The velocity measurements were performed with a noncommercial heterodyne laser interferometer which is combined with a visual microscope (Khanna *et al.*, 1996). The focused laser beam coincides with the center of the image of the microscope displayed on a video monitor; the microscope allows visualization of the stapes, while the interferometer measures the vibration at the selected location in response to sound applied to the ear.

For the velocity measurements the animal's body was resting on an object support, while its head was immobilized by firmly clamping the head holder, which was cemented to the exposed top surface of the skull, to a post affixed to the positioning system. The positioning device provides micrometer positioning (with read-out of the coordinates) in three orthogonal directions (x,y,z) and allows changing the observation direction by means of two large, nested goniometers that are precisely aligned to rotate the object about the point of focus of the laser beam. (For details on this alignment see Appendix A.) The accuracy on the x,y,z coordinates is on the order of $5-10$ μm. It was determined by repeatedly focusing the interferometer beam on a given point while moving far away from this point intermediately. This small error is a combined result of the high sectioning depth of the interferometer (the interferometer carrier is 3 dB down at a distance of ± 13 μm) and the good positioning reproducibility of the translational stages.

The sound source and sound driver were put back in the EC insert in a position closely equal to that during SV pressure measurement. With the goniometers in their midposition ($H=0°$, $V=0°$) the stapes was imaged, and we focused precisely on a chosen reference point (e.g., on the stapes head) and the coordinate read-out system was reset. For a set of

points (typically seven to eight) as widely spread as possible on the entire visible part of the stapes we recorded the velocity of the stapes in response to pure tone stimuli, generated with a constant voltage input to the driver (SPL ranging from 85 to 95 dB up to 30 kHz and dropping to a lowest values of 70 dB at 40 kHz). The frequency was stepped between 200 Hz and 1 kHz with a step of 200 Hz, between 1 and 10 kHz with 250 Hz, between 10 and 20 kHz with 500 Hz, and between 20 and 50 kHz with 1000 Hz. For further processing all velocities (amplitude and phase) were expressed relative to the simultaneously measured EC pressure.

Using only one goniometer at a time, the stapes was rotated maximally in positive and negative directions without losing its image in the microscope (which occurs when the bulla wall blocks the illumination and interferometer laser beam). At each extreme angular position new velocity measurements were recorded. The procedure was sped up by storing the coordinates of the observation points used at $H=0°$, $V=0°$ and refocusing on the points automatically, using stepper motors driving the x,y,z positioning system. Note that (i) the coordinate readings for a given observation point in focus did not change when the object was rotated, as the rotation center coincided with the focused point and (ii) that for calculation of 3D rigid body motion components it is not mandatory to measure at the same points for different observation angles; if an observation point was no longer accessible after rotation of the object it was replaced by another point. This freedom of changing the observation points when using a different observation direction comes in very handy and is a result of a modification we made to the procedure we used hitherto to determine the 3D components of the malleus (Decraemer and Khanna, 1996) and stapes (Decraemer and Khanna, 2000; Decraemer *et al.*, 2000).

It is also important to point out that during the rotation the animal, sound driver, and probe microphone moved together as they were all mounted to the same object holder post of the positioning system keeping relative positions unchanged.

F. Processing of the data

1. 3D velocity components in the experimental reference system

In earlier studies on malleus and stapes vibration in cat (Decraemer and Khanna, 1996, 2000; Decraemer *et al.*, 2000) we determined the 3D motion in two steps. First the 3D velocity components of at least three points on a given ossicle were calculated and in a second step these results were used to calculate the 3D rigid body motion of the ossicle. To calculate the 3D components for a given point we had to observe the same point from several observation angles. This constraint is now relaxed: different points can be observed. The coordinates of all observation points and the amplitude and phase of the velocity at these points for all (here five) observation directions were used in a single fit to calculate the rigid-body motion components directly. These values describe the motion as the translation of a reference point on the body and the rotation of the body about this point. We determined the complete 3D motion, and thus
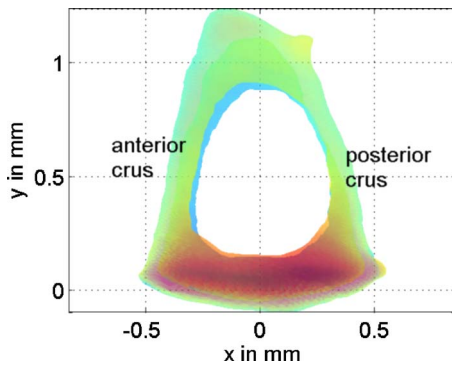
FIG. 2. (Color online) A 3D model of a gerbil stapes is brought in an intrinsic reference frame: The origin is at the footplate center, the x axis follows the long footplate axis in the posterior direction, the y axis is perpendicular to the footplate directed to the center of the stapes head, and the z axis is pointing out of the figure plane to the reader and follows the short axis in the inferior direction.

FIG. 3. (Color online) The model of the stapes is shown after registration in the position of the stapes during the experiment defined by the observation points (stars) and the extra set of "anatomical points" (circles). The insert at the top right shows a view of the stapes through the microscope during the experiment.

three components of both the translation and rotation vectors. Details of the mathematical procedure are given in Appendix B.

The assumption of rigid body behavior made during this calculation might seem questionable at first, the stapes of gerbil being such a thin and fragile structure. Its special design, however, with its two symmetrical widely spread crura with thin strongly curved walls and with its bowl-shaped footplate reinforced by an upstanding rim make it extremely light and sturdy at the same time (see the 3D stapes model in Fig. 2 or for a more detailed stapes model see Decraemer *et al.*, 2003).

Note that the motion components that we first obtain are expressed in the original coordinate system used during the experiment. In this reference system the object under study is quite arbitrarily oriented, which makes the straightforward interpretation of the frequency responses of the rotation and translation components almost impossible. In the next section we will explain how we can transform these components in a meaningful intrinsic coordinate system for the stapes.

2. 3D velocity components in an intrinsic reference system

As we are especially interested in the piston and the tilting components of the footplate we express the rigid-body motion components of the stapes in the following intrinsic reference frame: the origin at the center of the footplate, the x axis pointing in the posterior direction along the long axis of the footplate, the y axis perpendicular to the footplate pointing to the center of the stapes head, and the z axis running along the short axis of the footplate, completing a right-handed reference system (Fig. 2).

To calculate the motion components in the intrinsic reference system we first determined the coordinate transform that rotates the stapes from its experimental position into a position parallel to its position in the intrinsic frame and then shifted the stapes to let the origin coincide with the footplate center. We determine this transform in an intermediate step using a 3D stapes model as described in the following (Sec. II F 2 a). The coordinate transform that brings the stapes model from the experimental into the intrinsic frame will
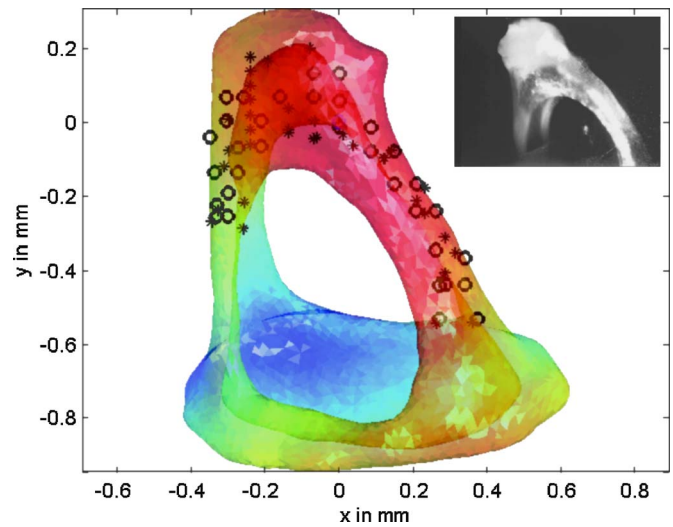
also transform the translation and rotation vectors in the intrinsic reference frame as explained in Appendix C.

a. Constructing and registering a 3D stapes model. Based on a micro-CT (CT: computed tomography) scan of the excised experimental ear we constructed a 3D model of the stapes. This model is defined in the reference system of the CT scanner which again is different from the experimental reference system as it depends on the position and orientation of the temporal bone during the scan. In a first step the model was registered in the reference frame that was used during the experiment and in which the coordinates of observation points and stapes vibration were recorded. The registration was done with an iterative closest point procedure described by Besl and McKay (1992). The number of observation points at which velocity was measured (less than 20) was too small to define the position of the stapes during the experiment with sufficient precision to make an accurate registration. To enhance registration, during the motion experiment we measured the coordinates of an extra set of about 50 "anatomical" points. This was a fast measurement that consisted of focusing the laser of the interferometer at points spread across the entire visible part of the stapes and automatically recording the coordinates (without vibration measurements). The procedures to make the 3D model and to register the model were described in a previous paper (Decraemer *et al.*, 2003). The 3D model registered in the set of experimental (stars) and anatomical points (circles), along with a picture of the stapes taken during the experiment through the microscope (insert) are shown in Fig. 3.

b. Coordinate transform from experimental to intrinsic frame. In a second step the model in the experimental reference frame was iteratively rotated about the x, y, and z axes until it was in a position parallel to its intrinsic position. In this way we obtained the rotation matrix of the required coordinate transformation. The vector from the experimental origin to the center of the footplate (new origin) determines the origin shift, completing the coordinate transform. The stapes in its intrinsic reference frame is shown in Fig. 2.
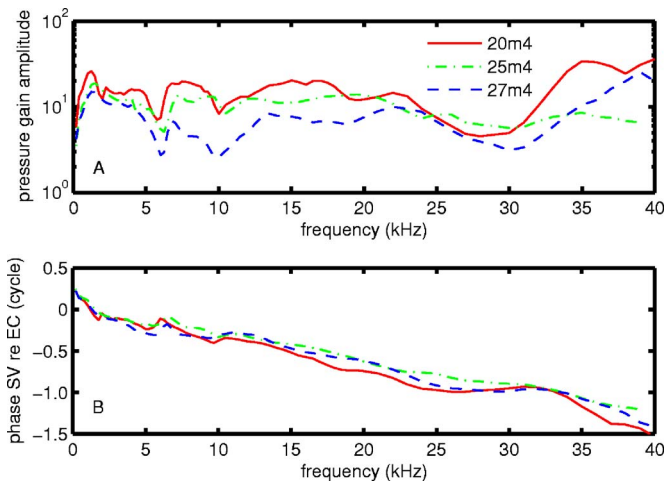
FIG. 4. (Color online) Pressure in the scala vestibuli relative to EC pressure ("middle ear pressure gain") is shown for three animals, 20m4, 25m4, and 27m4 (see legend). The amplitude of the gain (upper panel) is smooth with frequency, and varies in a small band around a level of ~10 (20 dB). The phases of the gain (lower panel) for all three animals were similar; the three curves start at +1/4 cycle at 190 Hz, rapidly decrease to 0 cycle at about 1.5 kHz where we observe a small peak, and decrease further to about 6 kHz where a second small peak is seen. From then the three phase curves show a gradual and approximately linear decrease with frequency to a total phase lag of about −1.5 cycle at 40 kHz.

G. Relative phase between pressure sensor and interferometer

In order to couple the measured pressures and velocities, we need to know the difference in the processing between the pressure sensor and the interferometer, in particular the relative phases of the two systems. The pressure sensor is composed of a gold-coated thin film membrane, whose position is interrogated by light emitting diode (LED) light via a fiber-optic lever. An optic lever is a simple displacement sensor, whose output is a voltage whose variations are proportional to displacement. To determine the relative phase between the interferometer and sensor we focused the interferometer laser on the sensor membrane. We drove the membrane with a free-field harmonic sound source while simultaneously monitoring its velocity with the interferometer on one side and its displacement with the sensor's optic lever on the other side. To compare amplitudes, we divided the reported velocity by ω. The displacement versus frequency measured with the interferometer and pressure sensor optics had very similar shapes in amplitude, and a relative phase that was a nearly straight line, indicative of a delay. The interferometer's demodulation electronics (a commercial REVOX FM demodulator) are based on a delay line that we knew produced a delay of ~15 μs (Willemin *et al.*, 1988). In our data, the phase-frequency data were well fit by a straight line with 14.1 μs delay, and this correction was used to compare interferometer and pressure sensor outputs.

III. RESULTS

A. Scala vestibuli pressure

Pressure in the SV relative to EC pressure is shown for three animals, 20m4, 25m4, and 27m4, in Fig. 4. These curves represent the pressure gain produced in the middle ear: the gain amplitude [Fig. 4(a)] was a smooth function of frequency, fluctuating in a small band around a level of 10 (20.0 dB) with extreme values of 3 (9.5 dB) and 20 (26.0 dB). In the low frequency tail the three animals showed a fast increase in gain with frequency resulting in a small maximum at ~1.5 kHz, followed by a steady decrease producing a minimum at about 6 kHz. The gain was again somewhat higher in a short frequency interval ending with a small dip at 10 kHz. For animal 25m4 (dashed dotted line) the gain was remarkably constant all the way up to the highest frequencies measured. Animals 20m4 and 27m4 had a broad and shallow minimum at 30 kHz, followed by a rise to their overall highest values.

Phase for all three animals was very similar [Fig. 4(b)]: the three curves started at +0.25 cycle at 190 Hz, rapidly decreased to 0 cycle at about 1.5 kHz (here the amplitude showed a small peak), then decreased further to about 6 kHz where a small peak appeared (here the amplitudes produced a dip). From then on the phases for all animals decreased linearly with frequency and a total phase lag of about −1.5 cycle had accumulated by 40 kHz. These results, particularly the flat magnitude and linear phase were shown in Olson (1998) and Dong and Olson (2006). Some systematic reasons for variations in results in different experiments, particularly in gain, were discussed in detail in Dong and Olson (2006).

B. Three-dimensional stapes velocity

1. Measurement accuracy based on repeatability: Comparison of changes due to observation angle with changes developing in time

In order to determine the 3D rigid body motion of the stapes we must perform repeated vibration measurements from different observation angles. Optimally we would like to measure along three orthogonal directions and spread the observation points over the entire ossicle. The anatomy of the middle ear does not provide such access for uniaxial interferometric measurements; only a limited part of the stapes is visible (Fig. 1) and variations in the observation angle are typically limited to ±15° to 20°. The changes in vibration for these relatively small angle variations are consequently also small. High measuring sensitivity is required and we must minimize changes with time by, e.g., letting the temperature stabilize, preventing desiccation, and by reducing total recording time. Figure 5 shows first the repeatability of the frequency responses for amplitude [Fig. 5(a)] and phase [Fig. 5(b)] of the displacement normalized to sound pressure level at a single point recorded with the goniometers set at $H=0°$ and $V=0°$. Figure 5 shows the mean and standard deviation for four runs (three repeats recorded within 7 min and one repeat measurement made 40 min later as a control measurement in the further course of the experiment), at a given reference point that was chosen at a retraceable detail in the image, e.g., at the middle of the stapes head. Figure 5 shows also the variation in response to changes in observation angle ($H=0°$, −15°, and +12°, and all $V=0°$) at a second observation point, halfway down on the posterior crus. We see that at most frequencies the change

(a)

(b)
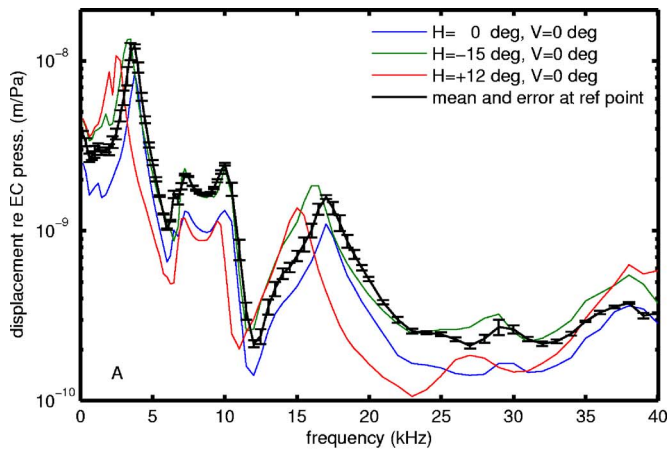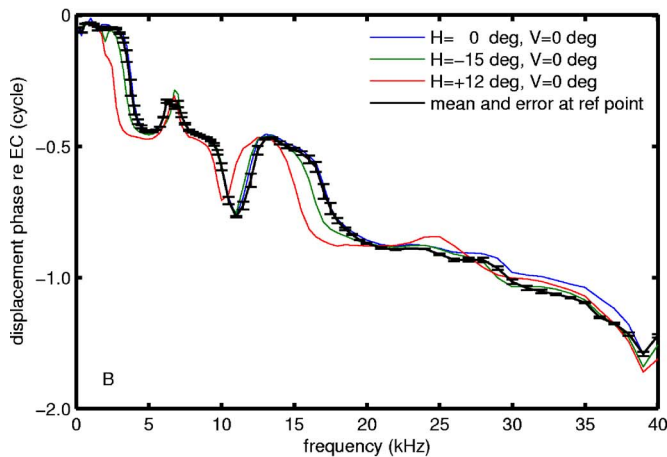
FIG. 5. (Color online) Repeatability of frequency responses for the amplitude (A) and the phase (B) of the displacement (per EC pressure) at a single point is shown by plotting the mean and its standard deviation for four runs (three repeats recorded in a time window of 7 min, and one control, 40 min after the last run, all with $H=0$ and $V=0$). At a different position, measurements were performed for three observation angles. The variation in response with observation angle at this point is also shown by plotting frequency responses for the different viewing angles (see the legend). The spread in change with observation angle for both amplitude and phase is at most frequencies substantially larger than the measurement uncertainty.

with observation angle of both amplitude and phase was substantially larger than the measurement uncertainty (here quantified by the repeatability of the measurements), as required. As we show results for two different observation points one can also judge the changes that are observed due to a change of the observation point by comparing the plot at $H=0°$, $V=0°$ with the curve showing the mean for the four runs at the same angle.

2. Rotation and translation frequency responses of the stapes in the experimental reference frame

Rigid body motion is described by three translation and three rotation components. In this and the next section we will show motion as rotational and translational displacement per unit pressure. These results were found using the methods of Appendix B operating on data like that in Fig. 5. Later, some of the displacement components will be converted into velocity and correlated with SV pressure in order
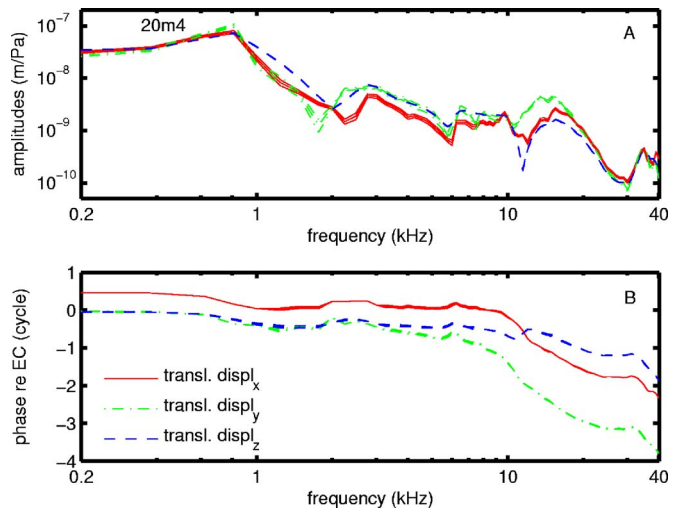
FIG. 6. (Color online) Frequency response for the three components of the translation displacement per EC sound pressure, expressed in the experimental reference frame. The experimental error, determined using a Monte Carlo method, is shown by plotting the curves with the component value + error and component value − error. The errors are small compared with the overall variations with frequency; translation displacement amplitudes (A) and phases (B) were determined with good accuracy.

to find acoustic impedance. We present results for the stapes of one gerbil (20m4). Figure 6 shows translation, Fig. 7 shows rotation. The components were calculated in the reference frame of the experiment. The different line types for the x, y, and z components are clarified in the legend. Due to the arbitrary nature of the experimental reference frame, the displacement components are difficult to interpret. Of major importance however is the question of how accurately these components can be determined. We therefore postpone the discussion of their significance until they are recalculated in the intrinsic frame and we restrict the discussion here to the error calculation.
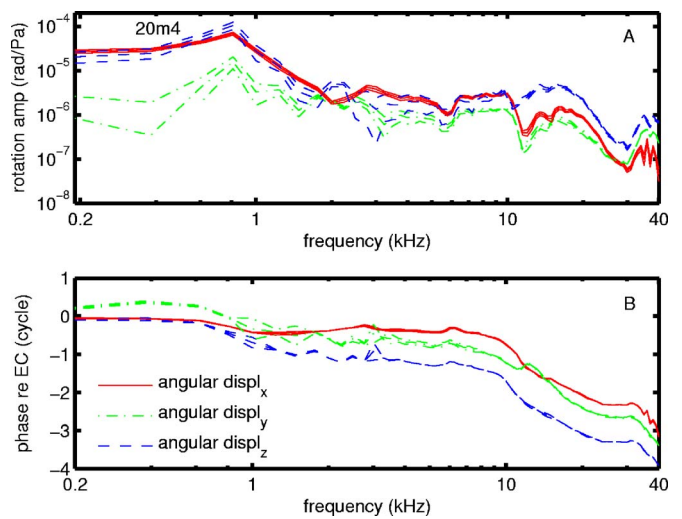
FIG. 7. (Color online) Frequency response for the three components of the rotation displacement per EC sound pressure, expressed in the experimental reference frame. The experimental error, determined using a Monte Carlo method, is shown by plotting curves with the component value + error and − error, as in Fig. 6. The errors on the rotational motion values are somewhat larger than those on the translation component values, but still show good accuracy.

The motion components were obtained with the method of Appendix B, using a least-squares fit with observations at all angles given the same weight. The estimate of the motion uncertainty (standard error) was obtained using a Monte Carlo method (Press *et al.*, 1989). To apply this technique we produced a large number (100) of new "experimental" data sets (frequency by frequency) by adding random noise drawn from a normal distribution (with a mean of zero and a variance chosen in agreement with the experimental uncertainty) to the observed amplitudes and phases and recalculated the fit for each set of data. The uncertainty in a displacement or rotation value was calculated as the standard deviation of all values obtained for this parameter. The variances used in this process to generate new data sets were the absolute error corresponding to a relative error of 0.05 in the amplitude and an absolute error of 0.1 rad in the phase of the measured vibration. The variance values were based on averages over all stimulus frequencies of the standard deviations in amplitude and phase for repeated observations at the same location, some measured in direct succession, some after intermediate measurements at other locations as explained in Sec. III B 1. The errors found using the Monte Carlo data to derive angular and translation displacements are indicated in Figs. 6 and 7 by plotting the nominal displacement values (without added noise) plus and minus their statistical error. The logarithmic scale for the amplitudes and the large span of the phase data tend to make the errors look small (some traces nearly or completely overlap), but nevertheless it is obvious that the obtained accuracy was good; the mostly smooth up and downward going trend in the curves is genuine and not due to noise in the observations as this would have generated jagged frequency responses. Errors on rotation components tend to be larger than errors on translation components. The errors that were obtained using the Monte Carlo method are in perfect accordance with errors that were obtained from a weighted fit procedure—an observation is given more weight when its experimental error is smaller—which simultaneously produced errors on the rigid body velocities as elements of the covariance matrix (Press *et al.*, 1989). The weights were also chosen on the basis of a constant relative error on the amplitude and constant absolute error on the phase and were then converted into weights on the in-phase and out-of-phase components (cf. Appendix B). We obtained frequency curves that showed the same overall trend as those from the unweighted fit, but the curves showed a few sudden spikes, most probably caused by strong fluctuations of the in- and out-of-phase components as they are calculated as the amplitude times the sine and cosine of the phase angle. For this reason, for its intuitive simplicity and in view of its ease in transforming the errors in another reference system (Sec. III B 3) the Monte Carlo error calculation method was finally retained.

The goodness-of-fit was also estimated by the calculation of the mean deviation between the experimental amplitude and phase and their fitted counterparts. For the amplitude we obtained a typical mean relative deviation between 10% and 20%, for the phase a mean absolute deviation typically between 5° and 15°. We think that these are very acceptable values, seen that for a given frequency the ampli-
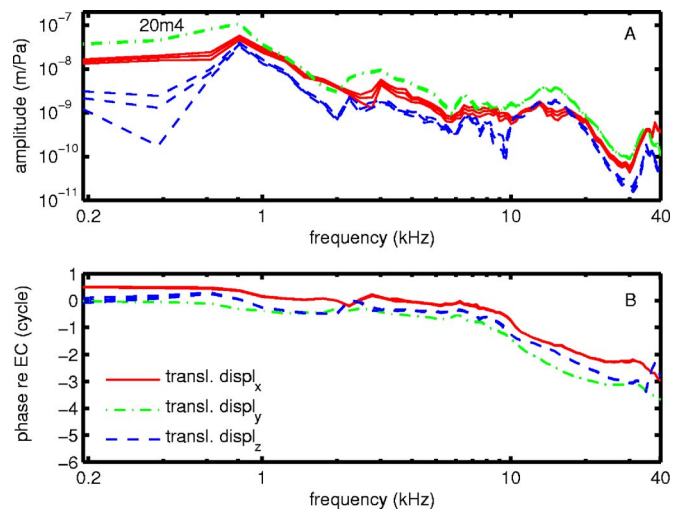
FIG. 8. (Color online) Frequency response for the translation displacement components per EC sound pressure for animal 20m4 expressed in the intrinsic reference system as defined in Fig. 2. (A) Amplitudes and (B) phases. Uncertainties are shown by plotting component + error and component − error curves. We see that peaks and valleys are genuine and that the piston component (dashed-dotted line) is most accurately determined (smaller error).

tudes of all observation points at all observation angles typically change within a factor of 3–10. Apart from the experimental uncertainty in the direct measurement, some of these deviations are caused by changes with time of the animal preparation that cannot be avoided when performing an extended experiment.

Let us mention at this time that the good quality of the rigid body model fit to the observed motion is indicating that the assumption of rigid body behavior is acceptable. This can be contrasted with the previous measurements of some of the authors (Decraemer *et al.*, 1994b) of manubrium motion. Unlike the compact stapes, the manubrium has an extended beam-like geometry. In that case, experiments clearly showed the existence of a bending mode on the thinnest part of the manubrium close to the umbo. In experiments to describe the 3D ossicular motion (Decraemer and Khanna, 1999, 2000), no observation points at this location were used. It is most probable that a structure such as a stapes will also have bending modes, but our results suggest that they add only to a small extent to the total velocity at a given point.

3. Rotation and translation frequency responses of the stapes in the intrinsic reference frame

To interpret the motion components they were transformed into the intrinsic frame of reference for the stapes as defined in Fig. 2. We have displayed components for three animals, 20m4 (translation: Fig. 8 and rotation: Fig. 9), 25m4 (translation: Fig. 10 and rotation: Fig. 11), and 27m4 (translation: Fig. 12 and rotation: Fig. 13), with amplitude (displacement normalized to EC pressure) in the upper panels (A) and phase (relative to EC phase) in the lower panels (B), to illustrate interanimal similarities and dissimilarities.

Calculation of the uncertainties in the results in the new frame with error propagation formulas is not straightforward.
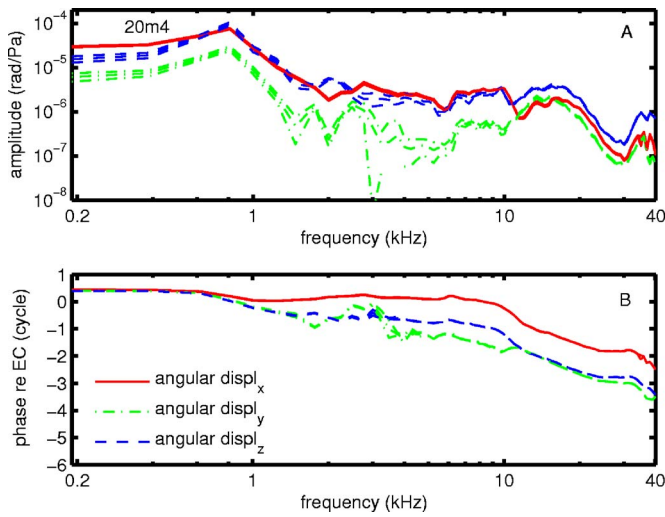
FIG. 9. (Color online) Frequency response for the rotation displacement components per EC sound pressure for animal 20m4 expressed in the intrinsic reference system as defined in Fig. 2. (A) Amplitudes and (B) phases. Uncertainties are shown by plotting component value + error and value − error curves.

FIG. 11. (Color online) Frequency response for the rotation displacement components per EC sound pressure for animal 25m4 expressed in the intrinsic reference system as defined in Fig. 2. (A) Amplitudes and (B) phases. Uncertainties are shown by plotting component value + error and value − error curves.

The Monte Carlo technique can also be used to simplify this calculation. To produce random sets of rigid body motion components in the experimental reference frame we added to the fitted components Gaussian noise with a variance equal to $\frac{3}{4}$ of the standard errors (variances equal to the standard errors yielded some unrealistically large deviations when extreme values in the tails of the distribution were drawn). All these synthetic displacement components were converted in the intrinsic frame and the error on a given component was obtained as its standard deviation. The uncertainties are indicated with lines above and below the curves (parameter ± error) for each displacement component. As noted earlier, we see that peaks and valleys are genuine and not an artifact of the analysis.

a. Translation Plots of the frequency responses for the translation displacement amplitude per EC sound pressure for the three animals [Figs. 8(a), 10(a), and 12(a)] reveal similar features for corresponding components, with absolute levels of vibration differing about three to five times depending on the component up to 20 kHz and up to 10 times at the highest frequencies. Overall all translation components were of similar size, with the piston component (*y* axis) usually largest. The piston component was high and quite constant at low frequencies, peaked at about 1 to 2 kHz, decreased between 4 and 8 kHz, had a dip at about 5.8 kHz and a small peak at 8 kHz [Fig. 8(a)] or a dip at 6 kHz and a small peak at 7 kHz [Figs. 10(a) and 12(a)] and a small or more pronounced peak at 10 kHz in Figs. 8(a) and 10(a), respectively. Then it increased slowly to a broad maximum at about

FIG. 10. (Color online) Frequency response for the translation displacement components per EC sound pressure for animal 25m4 expressed in the intrinsic reference system as defined in Fig. 2. (A) Amplitudes and (B) phases. Uncertainties are shown by plotting component + error and component − error curves. We see that peaks and valleys are genuine and that the piston component (dashed-dotted line) is most accurately determined (smaller error).

FIG. 12. (Color online) Frequency response for the translation displacement components per EC sound pressure for animal 27m4 expressed in the intrinsic reference system as defined in Fig. 2. (A) Amplitudes and (B) phases. Uncertainties are shown by plotting component value + error and value − error curves. We see that peaks and valleys are genuine and that the piston component (dashed-dotted line) is most accurately determined
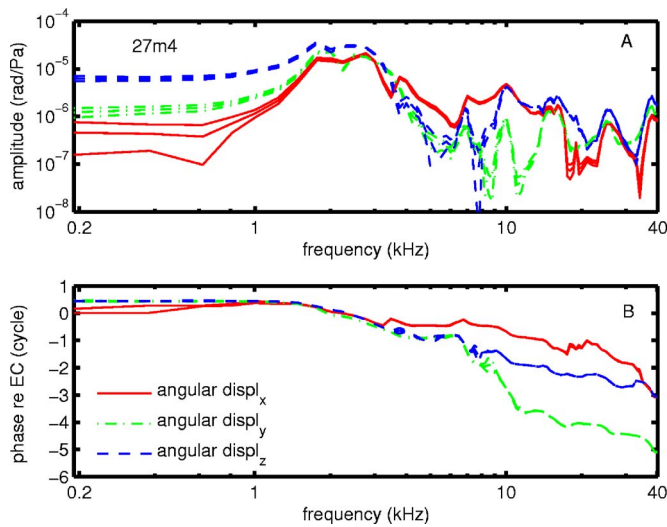
FIG. 13. (Color online) Frequency for the rotation displacement components per EC sound pressure for animal 27m4 expressed in the intrinsic reference system as defined in Fig. 2. (A) Amplitudes and (B) phases. Uncertainties are shown by plotting component value +/− error curves.

FIG. 14. (Color online) Amplitude (A) and phase (B) of the pressure gain (SV pressure relative to EC pressure, thick solid line) as a function of frequency for animal 20m4. The piston component velocity (thin solid line) and angular velocities about the long axis (dashed-dotted line) and short axis (dashed line) of the footplate are plotted as well to correlate pressure and velocity. To plot the amplitudes of the velocities on the graph in the scale of the gain, the piston velocity is expressed in units of 10^{-5} (m/s)/Pa, and the angular velocities in 10^{-3} (rad/s)/Pa.

15 kHz [Figs. 8(a) and 12(a)] or 18 kHz [Fig. 10(a)] followed by a decrease, fast in Fig. 8(a), slow in Figs. 10(a) and 12(a), to a minimum at 30 kHz. For higher frequencies it rose again to a maximum at 36 kHz of different height according to the animal. The in-plane translation along the long footplate axis (x) started out three times smaller that the piston component, but rose rapidly to become equal or even bigger than the piston component in a small frequency band around 2 kHz [Fig. 8(a)] or 3 kHz [Figs. 10(a) and 12(a)]. Then it decreased between 4 and 8 kHz but displayed peaks at the same frequencies as the y component. In Fig. 8(a) the x component peaked at 13 kHz at a value of half the y component and in Figs. 10(a) and 12(a) at 15 kHz to values slightly higher to equally high as the y component. The in-plane translation along the short footplate axis (z) had overall the lowest amplitude and the largest uncertainty. It had, as the other components, its highest amplitude at low frequencies and then dropped considerably. At 7 kHz [Fig. 8(a)], respectively, at 10 kHz [Figs. 10(a) and 12(a)] it showed a peak, also present in the corresponding y and z components, followed by a sharp dip. At around 15 kHz it rose again to a broad plateau. Two animals [Figs. 10(a) and 12(a)] showed a slight to more pronounced maximum at 28 and 25 kHz, respectively. After a dip at about 30 kHz the z components finally followed the same increasing trend to exhibit a maximum at around 37 kHz.

The phase of the piston component (y) for the three animals started out in-phase with the EC pressure at low frequencies. The phase remains zero up to the first peak in amplitude where a lag of about 0.25 cycle was observed [small step in Figs. 8(b) and Fig. 12(b), more gradual change in Fig. 10(b)], in agreement with a stiffness controlled system. In Fig. 8(b) the phase lag increased slowly and gradually below 10 kHz and much more rapidly above 10 kHz to reach about −3.5 cycle at 40 kHz. Phase lag accumulated with frequency in a very gradual way over the entire frequency range in Fig. 12(b). In Fig. 10(b) the phase took two steps-down of ∼0.5 cycle in small frequency bands, accompanied by strong changes in amplitude [Fig. 10(a)] so that it ended up with larger total lag compared to Fig. 12(b).
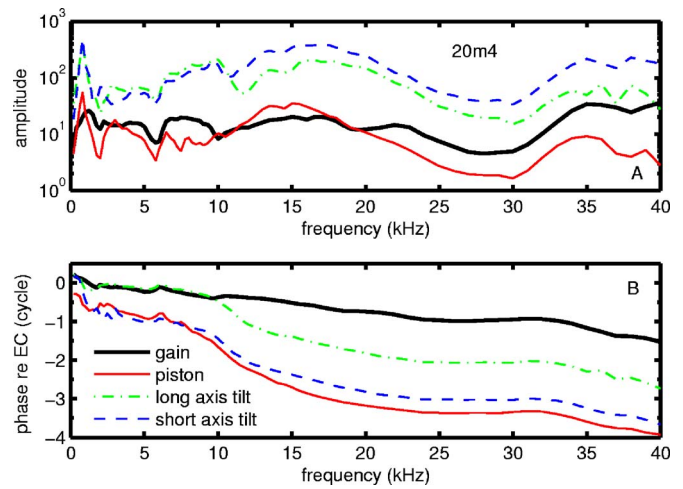
Phases for the long footplate axis components (x) started with a lead of 0.5 cycle re pressure [Figs. 8(b), 10(b), and 12(b)]. The short footplate axis component (z) once led by 0.5 cycle [Fig. 10(b)] and twice started in-phase with pressure [Figs. 8(b) and 12(b)]. In Fig. 8(b) all phases follow roughly the same trend. Phases for y and z were quite similar in Fig. 10(b), in Fig. 12(b) this holds well for the x and y components.

b. Rotation. The angular displacements complementing the translation displacements of Figs. 8, 10, and 12 are given in Figs. 9, 11, and 13, respectively. The amplitudes (in rad/Pa) are shown in the upper panels (A) and the phases relative to EC in cycles in the lower panels (B). In Figs. 9(a), 11(a), and 13(a) the three curves are intertwined In Fig. 13(a) the angular displacement components differ more from one another than in Figs. 9(a) and 11(a). Generally the rotation amplitudes were at their highest around 1–3 kHz where also the respective translational amplitude peaked, exhibited peaks and valleys while decreasing to their lowest values at about 30 kHz in Fig. 9(a), 23 kHz in Fig. 11(a), and 35 kHz in Fig. 13(a) and then climbed again to a maximum at around 37 kHz. The peak seen in all translation amplitude components at 10 kHz was also present in all rotation amplitude components. The x-component phase started off twice at 0 [Figs. 11(b) and 13(b)] and once at 0.5 lead [Fig. 9(b)], all the y- and z- components at 0.5 cycle phase lead relative to EC pressure. The three animals showed no common trend in the way the rotation phase lag accumulated with frequency.

C. SV pressure versus piston velocity and footplate tilt

We calculated the stapes piston velocity per EC pressure for animals 20m4, 25m4, and 27m4 (data were shown previously as displacements in Figs. 8–13) and we compare these velocities (y component of the translation velocity, thin solid line) with the individual pressure gains (thick solid line) in Fig. 14 (20m4), Fig. 15 (25m4), and Fig. 16 (27m4).

Decraemer *et al.*: Intracochlear pressure and 3D stapes velocity
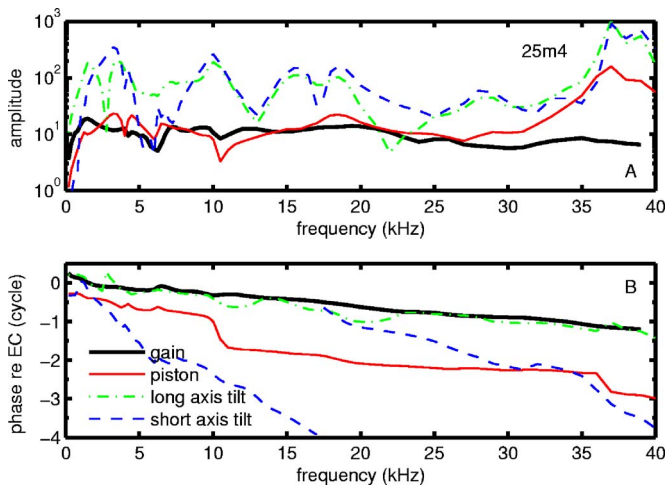
FIG. 15. (Color online) Amplitude (A) and phase (B) of the pressure gain (SV pressure relative to EC pressure, thick solid line) as a function of frequency for animal 25m4. The piston component velocity (thin solid line) and angular velocities about the long axis (dashed-dotted line) and short axis (dashed line) of the footplate are plotted as well to correlate pressure and velocity. To plot the amplitudes of the velocities on the graph in the scale of the gain, the piston velocity is expressed in units of 10^{-5} (m/s)/Pa, and the angular velocities in 10^{-3} (rad/s)/Pa. The discontinuity in the dashed phase curve is intentionally introduced by incrementing the phase angle with three cycles above ~17 kHz to enhance interpretation of the other curves on the graph (B).

To plot pressure and velocity on the same scale we multiplied the velocity by a scaling factor mentioned in the figure caption. To facilitate comparison over the entire frequency range we now use a linear frequency scale. It is clear that the pressure gain amplitudes [Figs. 14(a), 15(a), and 16(a)] were smooth functions of frequency. The piston component velocities relative to EC pressure amplitude were also relatively smooth, although less-so than pressure. In contrast, the footplate tilt showed considerable frequency structure. The gross trend of the curves for all three animals is fairly com-
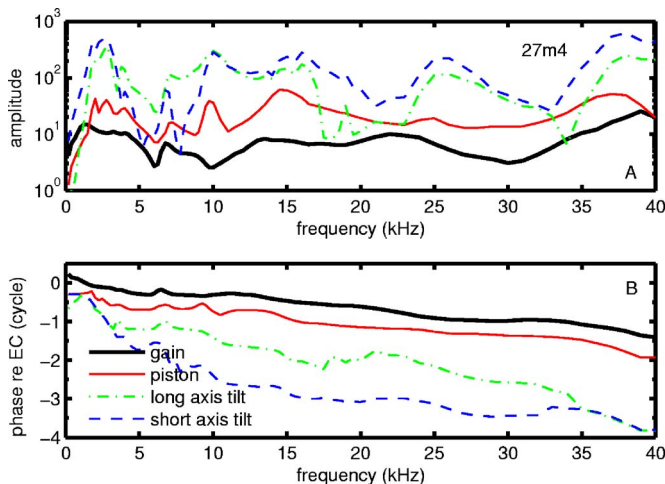
FIG. 16. (Color online) Amplitude (A) and phase (B) of the pressure gain (SV pressure relative to EC pressure, thick solid line) as a function of frequency for animal 27m4. The piston component velocity (thin solid line) and angular velocities about the long axis (dashed-dotted line) and short axis (dashed line) of the footplate are plotted as well to correlate pressure and velocity. To plot the amplitudes of the velocities on the graph in the scale of the gain, the piston velocity is expressed in units of 10^{-5} (m/s)/Pa, and the angular velocities in 10^{-3} (rad/s)/Pa.

parable. Differences between pressure and piston velocity were at the very lowest frequencies where the growth of the piston velocity with frequency was seen to be slower than the pressure gain growth with frequency. Some clear peaks in piston velocity amplitude, but of overall change in amplitude limited to ±66 dB, were not seen in the pressure (e.g., 25m4 at 3.5, 4.5, 18, and 37 kHz or 27m4 at 10 and 15 kHz). Some dips in the piston amplitude were accompanied by dips in the pressure (20m4: 5 and 30 kHz; 25m4: 6 and 10 kHz; 27m4: 6 kHz), but sometimes this was not true and a peak in velocity amplitude could even go along with a dip in pressure (27m4: 10 kHz). Disregarding these small wiggles the piston velocity was quite flat in a wide frequency band from 2.5 to 32.5 kHz for the three animals and the amplitude stayed high up to the highest frequencies measured. The pressure gain phase for all three animals [Figs. 14(b), 15(b), and 16(b)] winds approximately linearly down to a lag of about −1.5 cycle at 40 kHz (as already shown Fig. 4); the piston velocity phase in one case had a very similar excursion (27m4, Fig. 16) and in the others was more complicated.

Pressure in the inner ear is produced when the footplate compresses the fluid in the cochlea and therefore only displacements perpendicular to the footplate are expected to induce pressure. The motion components that give rise to such displacement are the piston component of the translation and the tilt of the footplate about its short and long axes. At first it seems reasonable to exclude the tilting from producing a pressure that can propagate further down the cochlea since the tilting motion around the origin at the footplate center does not result in a compression, merely a redistribution of the fluid at the stapes. Let us contrarily hypothesize that such rotations are also effective in sound transmission. Can we find evidence in our data to support this viewpoint? In the above, the piston velocity was less smooth than the SV pressure, and differed from the pressure in the details of the frequency response. We hypothesize therefore that stapes rotation fine tunes the transmission of sound by the ossicular chain in order to produce the smooth pressure response. To address this question we added to the plots of Figs. 14–16 the tilting components about the long (x axis, dashed dotted) and short (z axis, dashed line) footplate axes. The angular velocities were also scaled by a multiplier (see figure caption) so that again the same scale on the ordinate could be used. Angular components must be multiplied by the perpendicular distance to the concerned rotation axis in order to obtain the extra velocity in the piston direction that must be summed with the translation piston velocity; at the footplate center the extra velocity is zero and, e.g., for a point at the end of the long axis the extra velocity is half the length of the long axis (5×10^{-4} m) multiplied by the angular velocity (typically 10^{-1} rad/s) yielding 5×10^{-5} m/s, which is half the average translation piston velocity (piston velocity $\sim 10^{-4}$ m/s). Angular displacement therefore contributes significantly to the velocity of individual points on the footplate. Footplate tilt is frequency dependent (causing vibration mode of the stapes to change with frequency) and if it affects the SV pressure it can be expected it will do so at frequencies with large tilt.

To explore the hypothesis that tilting modes contribute
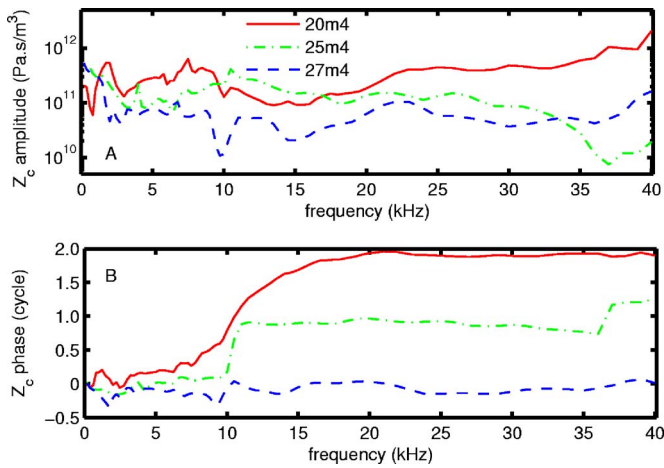
FIG. 17. (Color online) Cochlear input impedance Z_c [(A) amplitude; (B) phase] for all three animals (20m4, 25m4, and 27m4) as a function of frequency. The overall amplitude change over the very wide experimental frequency interval (200 Hz–40 kHz) remains small and this holds also for the phase if considered modulo 1 cycle.

to the pressure generation and play a role in smoothing the pressure gain versus frequency curve, we should look for frequency bands where the piston amplitude is small and where this is compensated by high angular amplitude. The effect does not have to be large as there were no frequencies where the piston amplitude went through a deep minimum. In fact, dips in piston amplitude were not systematically accompanied by peaks in tilt amplitude. A deep minimum in piston amplitude, not reflected in the pressure and accompanied by a peak in the tilt amplitude, would have supported the hypothesis, but it did not occur.

The data do not show any systematic relationship between footplate rotation and SV pressure, either in magnitude or phase. The fractional variations in rotation are in general bigger than the fractional variations in piston-translation and the rotation components are less related to pressure than the piston component is. This is particularly evident in Fig. 15. Therefore, the piston component and the SV pressure are relatively well related and the differences that do exist cannot be explained by rocking motion.

D. Input impedance of the cochlea

With the present data we can calculate the acoustic input impedance of the cochlea as the ratio of the pressure measured in the SV fluid behind the footplate to the volume velocity of the footplate. Pressure and velocity measurements were measured within an interval of 1–3 h in the same animals. The volume velocity was calculated as the piston velocity times the area of the footplate projected in the x-z plane (measured on one of our 3D stapes models as 0.62 mm^2). Results for the three animals used in the previous section are shown in Fig. 17. The three impedance amplitude curves [Fig. 17(a)] are all in the same range, and relatively flat over the most of the frequency range, except at frequencies above 35 kHz. Phase angle curves [Fig. 17(b)] differ from each other, but a closer look shows that the angle is mostly equal to values close to 0, 1, or 2 cycle. The 10 kHz step in the 25m4 curve might be due to the unwrapping of

noisy data. The gradual and large phase excursion in the 20m4 curve is hard to explain, and needs to be repeated in truly simultaneous pressure and velocity measurements to warrant further interest.

IV. DISCUSSION

A. Scala vestibuli pressure and middle ear pressure gain

The present results on the SV pressure in gerbil are similar to results presented in previous papers of some of the present authors (Olson, 1998; Dong and Olson, 2006). Equipment and measuring procedure being the same this is to be expected and differences here are similar to the inter-animal differences found in those studies. Systematic sources of differences were discussed in Dong and Olson (2006).

In all three pressure gain curves shown (Fig. 4) we observed a pronounced dip in amplitude at about 6 kHz accompanied by a change in phase. We also found changes in vibration amplitude and phase occurring at the same frequencies (Figs. 8, 10, and 12). Similar changes in pressure and vibration were seen at comparable frequencies in gerbil experiments by Ravics *et al.* (1992), Teoh *et al.* (1997), and Rosowski *et al.* (1999). They showed that these were associated to resonances in a Hemholtz resonator formed by the bulla cavity and the hole made in the bulla wall, with the exact resonance frequency depending on the size of the hole.

B. Three-dimensional stapes motion

With our measuring technique we were able to determine the complete set of 3D rigid body motion components of the gerbil stapes, translational and rotational. In one experiment (20m4) almost the entire motion experiment was carried out while the animal was alive, the two other experiments were done in a just post mortem condition. We reported similar measurements for cat and human temporal bone previously (Decraemer and Khanna, 1997; Decraemer *et al.*, 2006).

The present measurements show larger deviations from piston motion than, e.g., our measurements on cat measured under similar experimental conditions enforcing velocity measurements with a large observation angle with the piston axis. To verify that this was not due to systematic error in our calculation we performed the following simulation experiment. We assumed that the piston performs a pure piston motion, with a frequency response as determined by one of our experiments (20m4). We calculated the amplitude and phase at the location of our experimental observation points for this motion in the experimental reference frame. We then applied Gaussian noise to these synthetic data (0.05 relative error on the amplitudes, 0.1 rad on the phase) and calculated the rigid body motion components for this data set in the intrinsic frame. We repeated this procedure 100 times and calculated the mean and standard deviations for the obtained motion components. The piston component obtained was perfectly in accordance with the piston component that we started out with. Translational motion components in the directions of the long and short axis were also produced but at all frequencies the amplitude of the component along the

Decraemer *et al.*: Intracochlear pressure and 3D stapes velocity

short axis was about 20 times smaller than the experimental amplitude and about 30 times smaller for the short axis component. The simulation also produced rotation components but again with amplitudes that at all frequencies were at least 15 times smaller than the experimental values. Noise on the experimental observations is therefore not at the origin of the nonpiston components that came out of our experiments, these components must be really present.

We do not know any other study that measured the full 3D motion of the stapes or any other middle ear ossicle in gerbil or any other species. In one study on stapes motion in human temporal bones with normal air conduction stimulation, Hato *et al.* (2003) measured the stapes footplate vibration up to 10 kHz at five positions marked by reflective microspheres, one in the center and the four others at the ends of the long and short axis. The laser Doppler velocimeter beam was aimed through an opening made in the facial recess at an angle of 50°–60° from the direction perpendicular to the stapes footplate. A cosine correction was used to correct for this angle offset in the calculation of the stapes footplate motion. As all measurements were done with a fixed observation angle, only three motion components, i.e., the piston motion and the tilt about the long and short footplate axis could be determined. They reported that at frequencies below 1 kHz the piston-like motion of the stapes footplate dominated, whereas at higher frequencies it moved with increasingly complexity, with rotational motion along both the long and short axis. They also stated that above 3 kHz the stapes moved predominantly in a rotary fashion, reaching a peak at 6 kHz; above 6 kHz the rotary motion decreased slightly compared with the piston-like motion. Observations on the not purely piston-like character of stapes motion in human temporal bones were also made by Heiland *et al.* (1999) and Ball *et al.* (1997) based on relatively qualitative observations of 3D motion in human temporal bones.

In gerbil we found that the piston motion does not dominate so explicitly the vibration mode at low frequencies but nevertheless we see a clear reduction in angular displacement amplitude below 2 kHz [animal 25m4, Fig. 11(a) and 27m4, Fig. 13(a)] for all components by factors varying between 0.02 and 0.01. Differences in the mode of stapes vibration among gerbil, cat, and human may be due to the specific anatomy and differences in the condition of the ears during the measurements; gerbil: very tiny middle ear measured in live or very fresh cadaver condition, cat: a few hours post mortem as we waited until the response at a given point had stabilized, and human: largest and most sturdy of the three ears, untreated temporal bone measured a few days post mortem. One striking anatomical difference in these examples resides in the footplate asymmetry. In cat and gerbil the asymmetry is quite small compared to that in human, where the footplate—and this is no pun—really resembles the shape of a footprint with the posterior end much broader than the anterior end Decraemer *et al.* (2003).

To put a figure on the mode of vibration of the stapes footplate we have calculated the displacement of the superior end point of the short axis as a result of rotation about the long axis and the displacement of the posterior end of the long axis as a result of rotation about the short axis. We
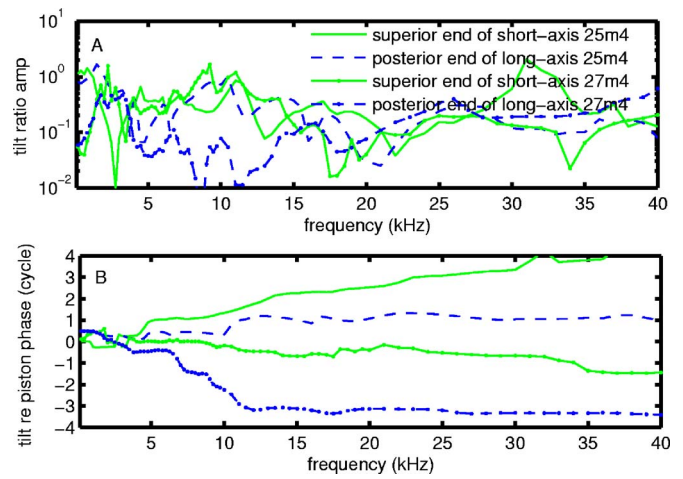
FIG. 18. (Color online) Comparison of displacement due to rotation about the long and short axis: the rotational displacement for the superior end of the short axis and the posterior end of the long axis are shown relative to the piston axis displacement. The amplitude (A) and phase (B) of this "tilt ratio" are shown as a function of frequency for animals 25m4 and 27m4.

plotted these displacements relative to the piston displacement for two animals (25m4 and 27m4) in Fig. 18. We see that for both animals the ratio of tilt to piston was the larger for the tilt about the short axis at almost all frequencies. Both "tilt ratio" amplitudes fluctuated with frequency, climbed to values greater than 1 at some frequencies, fell to values less than 1/100th at others, and were on average ~1/10. In certain frequency ranges tilts about short and long axis followed the same trend with frequency, but at other frequency ranges they were not correlated. The two animals showed strongly different tilt amplitudes. The plots of Fig. 18(b) show that the tilts about the long and short axes were not locked in phase, and that strong differences between the two animals were observed. We can conclude that in gerbil the piston component slightly dominates at frequencies below 2 kHz, but other components are also substantial and present throughout the entire frequency range.

Comments on the not purely piston-like character of stapes motion in gerbil were also made by Olson and Cooper (2000) based on a measurement of 3D motion of a single point on the posterior crus (by Cooper). In those studies, there appeared to be a transition from piston-like to more complex motion at ~6 kHz. Above this transition frequency, the motion was complex with substantial nonpiston motion, in accord with the present results.

C. Piston component measurement

Gerbil frequency responses of the stapes piston-velocity normalized by EC pressure were given by Rosowski *et al.* (1999) and Overstreet and Ruggero (2002). The velocity was measured using a unidirectional interferometer and an observation direction that made an angle with the piston axis because access for direct measurement of the stapes motion in line with the piston axis was not available. Rosowski *et al.* measured stapes motion up to 15 kHz, via 50 μm reflective plastic balls that were attached to the posterior crus via rubber cement and surface tension. The angle of incidence of the laser beam (Polytec laser vibrometer) was between 30° and

45° relative to the direction of the "piston-like" stapes translation, and a cosine correction was used to adjust the laser measurements to a value consistent with piston-like translation.

In the paper by Overstreet and Ruggero on high-frequency hearing sensitivity of the gerbil cochlea, stapes velocity was measured using a Dantec laser vibrometer (up to 40 kHz) focused on a reflective glass bead (diameter 20–30 μm) placed on the head of the stapes near the incudo-stapedial joint. They applied no correction to the measured velocity signal, arguing that the angle between the laser beam and the stapes vibration was always 30° or less.

The results of the two studies were very different. Rosowski et al. (1999) found stapes velocity features typical for a second-order spring-mass system, with a 6 dB/octave increase at low frequencies and a 6 dB/octave roll-off above 1.2 kHz. The phase varied from $+\pi/2$ to $-\pi/2$ between 50 Hz and 3 kHz, in accord with the spring-mass transition apparent in the amplitude.

Overstreet and Ruggero (2002, Fig. 2) show eight individual stapes velocity measurements as a function of frequency (500 Hz to 40 kHz) and their average. Each individual curve displays strong sudden fluctuations and between animals there are differences in level of up to 20 dB. The curves do not show a roll-off at higher frequencies, and the lack of high-frequency roll-off is emphasized in their results.

Our results agree and disagree with aspects of both studies. For animal 20m4 (Fig. 12) we saw a peak in piston velocity amplitude at about 1 kHz, with a phase transition in this frequency region of -0.5 cycle in agreement with Rosowski's study. While animals 25m4 (Fig. 15) and 27m4 (Fig. 16) showed a less pronounced local maximum near 1 kHz, the maximum in velocity amplitude occurred at about 2.5 kHz. The continued amplitude roll-off of Rosowski's curve is not seen in our measurements; we found a quite flat amplitude response all the way to the highest frequencies in accord with Overstreet. However, the dip in the frequency region <10 kHz could have been misinterpreted as the beginning of a roll-off if we had not extended our range of frequencies. The difference in frequency ranges explored in the different studies is likely the key to their different conclusions. In addition, calibration in the EC becomes somewhat ambiguous at high frequencies (above ~30 kHz in gerbil) and variations in calibration techniques can lead to systematic differences in results (Dong and Olson, 2006; Ravicz and Rosowski, 2004). Our average vibration levels were around 1×10^{-4} m/s Pa, which is in accord with the level of Rosowski (3×10^{-4} m/s Pa at the peak) and of Overstreet (average level ~3×10^{-4} m/s Pa).

Lets us point out further that the estimation of the stapes piston-velocity using an interferometer observation direction that makes a significant angle with the piston axis is prone to errors even when a cosine correction is applied (Decraemer et al., 2006). This will further account for differences between our results based on a true 3D measurement and results from single axis observations (see also Sec. IV D).

D. Do nonpiston components influence scala vestibuli pressure?

1. Correlation of pressure and motion components

A deep minimum in piston-like amplitude that was not reflected in the pressure would have been a clear cut demonstration that pressure can also be produced by tilt components: We did not see this in any of the three animals (nor in a few preliminary experiments).

The overall trends of the SV pressure and piston amplitude curves (both relative to EC pressure) are the same (Figs. 14–16): Over almost the entire frequency range (0.2–40 kHz) the pressure gain was steady, fluctuations occurred over wide frequency intervals and were not very pronounced (<20 dB). Smooth fluctuations in piston amplitude, not repeated in the pressure amplitude, were seen, but these were more a strong ripple on a curve than a trend change in the curve. Pronounced peaks and valleys on the tilt amplitudes were not systematically reflected as valleys and peaks in the piston velocity, or were they apparent in the pressure. These arguments all sustain the conclusion that only the piston component was effective in producing pressure in the SV. With the present measurements it is impossible to rule out completely that tilt motion contributes to SV pressure production, but such a contribution would have to be small. The findings are in agreement with the idea that it is the net volume velocity of the footplate (piston velocity times footplate area; see also the next section) that produces the pressure in the SV directly behind the footplate. The tilt must cause the fluid to slosh back and forth in a small volume behind the footplate but the related pressure does not propagate further inward.

To our knowledge there are no previous studies where in an intact ear of any species the SV pressure was measured in combination with the complete 3D motion of the stapes. The closest to our measurements were the simultaneous SV pressure and stapes motion measurements in studies on cochlear input impedance (Z_c) of cat (Lynch et al., 1982) and human temporal bone (Aibara et al., 2001). Lynch et al. measured piston motion along the piston axis using a Mössbauer technique, Aibara et al. estimated the piston component from unidirectional interferometer measurements at the center of the footplate with a viewing angle of 35°–50° with the stapes axis (these angles are not given in the paper by Aibara, but were taken from papers of members of the same laboratory, referenced in context with the stapes motion measurement (Kurokawa and Goode, 1995; Heiland et al., 1999)). Individual SV pressure and stapes piston motion curves are not shown in the Lynch et al. paper, therefore we will discuss it further in the section on impedance. Aibara et al. compare cochlear impedance magnitude with the stapes footplate rocking ratio found by Heiland et al. and concludes that a peak in the impedance curve at 6.5 kHz could be explained by a corresponding peak in the anterior/posterior rocking ratio curve, indicating that in this frequency region footplate motion is particularly complex and that a single point measurement at the center of the footplate is not adequate to determine the stapes piston motion. We would like to refine this statement: A measurement of center of footplate motion

under a large observation angle with the piston axis measures not only the cosine projection of the piston motion but it is also contaminated by the other translation components of stapes motion which, as we have shown, are present. However, truly rotation components about the footplate center will not contaminate the piston axis component estimate (Decraemer *et al.*, 2006) based on measurements at the center.

2. Footplate asymmetry and footplate volume velocity

The footplate (and its projection in a plane normal to the piston direction) is not flat but bowl shaped and it is asymmetric with respect to its long and short axes. Due to this asymmetry the tilt motion can produce a volume velocity that is not exactly zero (positive and negative part of the displaced volume not being equal). Earlier, we explored the hypothesis that tilt motion excites cochlear pressure experimentally; here we look at it theoretically, asking whether the tilt motion would in fact lead to net volume velocity into the cochlea at the stapes. In other words, what does a theoretical treatment have to say about the hypothesis that summing of the volume displacement due to piston and tilt acts as an integrating mechanism that fill in gaps of the piston volume displacement and smoothes the SV pressure?

The question could be answered by determining whether the volume velocity of the footplate is correctly calculated as the area of the footplate in the *x-z* plane times the piston velocity. We verified this by calculating directly the volume displaced by the footplate during its cyclic, 3D motion. The second calculation made no assumptions besides that the footplate is rigid. This was also one of the assumptions for the first calculation. As the footplate was not accessible during our experiments we have no real proof that it was indeed so. What we can say is that its bowl shape will make the footplate more rigid than if it were a flat plate. The footplate of the 3D stapes model can be described by a large number of small facets (small triangular, flat patch between three surface nodes). When the footplate is displaced a small volume (irregular pentahedron) is formed between the facet in its rest position and in its displaced position. We summed the volume for all the facets (954) of the footplate to obtain the displaced volume for the entire footplate. The calculation was repeated for 20 equally spaced time instants during a cycle. The obtained displacement as function of time within the cycle could be well fitted with a sinusoidal function (amplitude, phase). The volume displacement was converted to volume velocity. The calculation was done for all frequencies. The volume velocity obtained by integration was at all frequencies between 1% and 2% smaller than the simple calculation, *x-z* footplate area multiplied by piston velocity. The difference in phase was at all frequencies between −0.2° and +0.3°. Therefore, the simple multiplication is adequate and the asymmetry of the footplate does not give rise to an appreciable contribution to the footplate volume velocity caused by the tilt. In other words, the hypothesized integrating smoothing mechanism does not apply. This reinforces the experimental result, showing a lack of correlation between stapes rocking motion and cochlear pressure.

E. Cochlear input impedance

1. Gerbil

The only cochlear input impedance estimate for the gerbil is given by Overstreet and Ruggero (2002); it was obtained using the averaged stapes velocity measurement ($n = 8$, range: 1–40 kHz) of their own and SV pressure measurements taken from Olson (1998). For the magnitude their value ranged between 1 and 2×10^{11} Pa s/m^5 while the phase ranged between 1/8 and 3/8 cycle. These values for the magnitudes were in good agreement qua overall level, but the details of the curves were quite different. We have calculated the cochlear impedances on an animal per animal basis using pressure and velocity data collected in the same animal and found significant inter-animal differences; Overstreet and Ruggero (2002, Fig. 6) used an average velocity with large standard deviation (see the discussion on stapes motion) and pressure data reported from experiments in another laboratory (Olson, 1998) to calculate Z_c. Strong differences in the precise trend of their and our Z_c curves are to be expected. Phase angles of Z_c by Overstreet are limited within 0 and 0.5 cycle, while our values extend beyond 1 cycle. The averaging process tends to wipe out jumps due to phase unwrapping that could be present in our measurements. Phase values that fall outside −0.25 and +0.25 cycle are hard to believe as they suggest that the cochlea is producing net sound power. The fact that the pressure and motion measurements were not made simultaneously is one explanation for the observation. Alternatively, when notches and rapid phase variations occur in the velocity small errors in measurement could introduce full cycle phase excursions, for example in Fig. 15, following the conversion between reference frames.

2. Cat

The pioneering measurements on cochlear input impedance by Lynch *et al.* (1982, Fig. 15, frequency range: 10 Hz–6 kHz) were performed on cat, therefore we can only use them for an interspecies comparison. The middle ear was drastically altered by removing the tympanic membrane, malleus and incus and constructing a new middle cavity over the oval window and the stapes. This impedance curve was not as smooth as our curves, but as the accuracy on the stapes motion measurement was of the order of ±3 dB, and on the SV pressure of the ±5 dB, the fluctuations may have been caused by experimental error. The frequency range 0.01–6 kHz was also different from our present curves. We have no measurements in the low frequency range (20–200 Hz) where a pronounced increase in Z_c magnitude is seen by Lynch *et al.* (1982, Fig. 15). From 200 Hz on, phases by Lynch hover between −45° and +45°. If we regard our phases modulo 1 cycle the two data sets are in quite good agreement in the overlapping experimental frequency region.

3. Human

Experiments on human cochlear impedance by Aibara *et al.* (2001, Fig. 4, frequency range: 500 Hz to 10 kHz) show large fluctuations on individual curves and large interindividual variation in all of the presented measurements, EC

sound pressure to stapes footplate velocity, and EC sound pressure to SV sound pressure, resulting in a large variability in Z_c magnitude and phase. The average magnitude level reported for human is $\sim 2 \times 10^{10}$ Pa·s/m^3, which is about ten times lower than our present value for gerbil. Their average phase angle remains small (range 0 to $-60°$ or 0 to 1/6 cycle), which again is in agreement with our results, considered modulo 1 cycle.

V. CONCLUSION

We have presented measurements of the 3D velocity components of the stapes and have correlated these measurements with the SV pressure to verify whether nonpiston components of the footplate produce pressure behind the footplate and thus can give rise to an effective input to the cochlea and can help in realizing the remarkably smooth and flat pressure gain from ear canal to scala vestibuli as a function of frequency. We found no evidence for this hypothesis: Dips in the piston velocity were not systematically filled by peaks in tilt and conversely peaks in tilt were not apparent in the SV pressure. Further, we found that SV pressure and piston velocity correlated well overall but differences in trends of the curves were also present.

Although we collected pressure and velocity from the same individual animals, the measurements had to be made with a time interval between pressure and velocity measurements of at best 1 or 2 h, and sound source and EC microphone had to be taken out and replaced between the two experiment parts. During each experiment we experienced that pressure (SV and EC) and velocity measurements sometimes varied with time during the experiments, so differences between stapes piston velocity and SV pressure could be due to the fact that they were not simultaneously collected and to slight differences in the stimulus pressure. We therefore designed a new experimental protocol to measure piston velocity along a single axis and SV pressure at the same time. This ensures that the sound stimulus in the EC is exactly the same for both measurements. These experiments have been performed and will be reported in a follow-up paper.

APPENDIX A: ALIGNMENT OF THE CENTER OF ROTATION OF THE GONIOMETERS WITH THE FOCUSED LASER BEAM

Our setup is composed of an interferometer coupled to a confocal microscope (for a picture of the setup see Decrae-
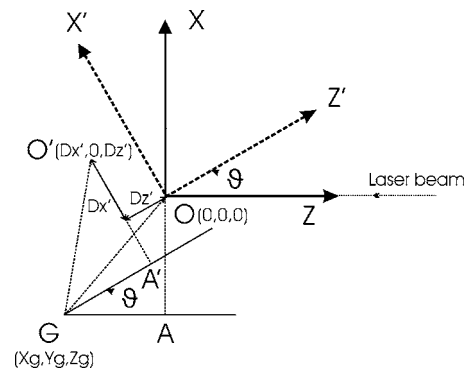
FIG. 19. O (0,0,0) is the origin of the sample's frame and corresponds to the point on the sample where the laser beam is focused. G is the position of the center of rotation of the goniometers. The distance between O and G is the "error" in alignment. After a rotation of θ about the y axis (points are rotated in the horizontal or $\hat{\mathbf{z}},\hat{\mathbf{x}}$ plane), point A becomes A', O becomes O' with $|\mathbf{AG}|=|\mathbf{A'G}|=|Z_g|$, $|\mathbf{OA}|=|\mathbf{O'A'}|=|X_g|$. With θ, X_g, and Z_g the correct base position can be calculated.

mer *et al.*, 1994a, Fig. 1; note that the axes in our actual experimental reference frame are inverted with respect to the situation shown in the insert: here the x axis is the same as in the insert (goes to the right), the y axis goes up, and the z axis toward the objective lens). The linear position of the object studied is controlled by a 3D linear positioning system. This positioning system can also be rotated around a vertical and a horizontal axis using a two goniometer system in a nested configuration. The outer horizontal goniometer rotates about a vertical axis (the y axis of our experimental reference frame), the inner goniometer about a horizontal axis. The goniometer base is adjusted to let the z axis of our experimental reference frame align with the interferometer observation direction (Fig. 19). In order to use the system optimally, the planes of focus of the microscope and the interferometer must be the same *and* the center of rotation of the goniometers must coincide precisely with the laser beam's focus. To this aim the goniometer base is mounted on crossed translational stages that can be repositioned in the x,z plane. The height (y axis) is adjusted by adding or removing thin metal sheets under the base of the goniometers. With the laser beam focused exactly at the center of rotation of the goniometers, the vibration of a given point can easily be studied from different angles, simply by rotating the goniometers. This appendix describes an efficient procedure for aligning the rotation center of the goniometers with the laser beam's focus.

As an object we use a very thin wire (10 μm diameter) that we first mount in a vertical direction in a ring shaped holder (diameter $\sim$10 mm). With both goniometers in their mid position (0°), the laser beam is focused on the wire. We let this point on the wire correspond to the origin O by resetting the linear coordinates to (0, 0, 0). The image of the wire and the position of the focused laser beam are traced on the screen of a video system with a marker.

Let us apply a rotation θ (θ corresponds to angle H in Appendix B) about the vertical axis (y). Looking down along the y axis in Fig. 19, we see that when G, the center of rotation of the goniometer, is not perfectly aligned with the focused laser beam, the point O rotates into a different point

O'. To bring this point back at the position of the laser focus we must move the object over a distance Dx' and Dz' that we can read off the linear scales.

The vector $\mathbf{OO'}$ can be expressed as a function of the coordinates of the center of rotation X_g, Z_g (vectors will be noted in bold; $\hat{\mathbf{x}}, \hat{\mathbf{z}}$ are unit vectors along x' and z' axes):

$$\mathbf{OO'} = Dx'\hat{\mathbf{x}}' + Dz'\hat{\mathbf{z}}',$$

(A1)

$$\mathbf{OO'} = \mathbf{OG} + \mathbf{GA'} + \mathbf{A'O'} = X_g\hat{\mathbf{x}} + Z_g\hat{\mathbf{z}} - Z_g\hat{\mathbf{z}}' - X_g\hat{\mathbf{x}}'.$$

After applying the coordinate transformation $\hat{\mathbf{z}} = \hat{\mathbf{z}}' \cos\theta - \hat{\mathbf{x}}' \sin\theta$ and $\hat{\mathbf{x}} = \hat{\mathbf{x}}' \cos\theta + \hat{\mathbf{z}}' \sin\theta$ we find

$$Dx' = X_g(\cos\theta - 1) - Z_g \sin\theta,$$

(A2)

$$Dz' = X_g \sin\theta + Z_g(\cos\theta - 1)$$

and

$$X_g = -\frac{Dx'}{2} + Dz'\frac{\sin\theta}{2(1 - \cos\theta)},$$

(A3)

$$Z_g = -Dx'\frac{\sin\theta}{2(1 - \cos\theta)} - \frac{Dz'}{2}.$$

Using these equations and the measured Dx' and Dz', distances X_g and Z_g can be calculated. To enhance accuracy the procedure is repeated for a number of increasing values of θ. The mean value of X_g and Z_g are calculated and used to adjust the position of the goniometer base in the x, z plane. The same procedure is performed with the wire mounted horizontally in order to determine Y_g and Z_g (Dy' and Dz' are measured for different V angles along the horizontal axis; we can reuse Eqs. (A2) and (A3) when we now let θ represent the angle $-V$). The height of the base of the goniometer is then adjusted following the calculated Y_g.

The centers of rotation of the vertical and horizontal goniometers were found to differ slightly in z direction (Z_g's obtained with V and H rotations differed by about 100 μm). The accuracy of the final alignment is therefore not perfect but for rotation less than 20° the displacement of the laser focus is less than 20 μm, and this was considered acceptable.

Remark that when X_g, Y_g, and Z_g are known for a not well aligned system we can calculate the observation point coordinates x, y, z in the reference system fixed to the interferometer, by subtracting Dx', Dy', and Dz' from, respectively, the coordinates x', y', z', recorded while the goniometers were turned.

APPENDIX B: CALCULATION OF THE 3D RIGID BODY VELOCITY COMPONENTS

To describe the stapes motion mathematically we use a stationary, inertial reference frame O_{xyz} attached to the interferometer. The origin of this fixed coordinate system is at the focus of the laser. In terms of the stapes, the origin is the point on the stapes that is in focus when the original vibration measurement was made (e.g., the zero position in Fig. 3). The stapes will be moved around and different points will be in focus when gathering the data to determine the motion, but the resulting description of 3D translational and rota-

tional motion will refer to this original stapes position. The coordinate system is right-handed, with x axis horizontal to the right, y axis vertical up, and z axis along the (horizontal) observation direction toward the observer. General motion of a rigid body may be decomposed in a rotation about the origin O followed by a global translation of the body. The vector velocity of a point P_i of the rigid body (stapes) can be written as (vectors are denoted in bold)

$$\mathbf{v_i}(t) = \mathbf{v_t}(t) + \mathbf{\Omega}(t) \times \mathbf{r_i},$$

(B1)

where $\mathbf{v_i}(t)$ is the velocity of P_i with components v_{ix}, v_{iy}, v_{iz} (this is the measured velocity); $\mathbf{v_t}(t)$ is the velocity of the global translational motion, with components v_{tx}, v_{ty}, v_{tz}; $\mathbf{\Omega}(t)$ is the angular velocity of the rigid body rotating about O, with components $\Omega_x, \Omega_y, \Omega_z$; $\mathbf{r_i}$ is the rest-position vector of point P_i with components x_i, y_i, z_i; and $\mathbf{\Omega}(t) \times \mathbf{r_i}$ is the velocity of the rotational motion component of P_i relative to O_{xyz}.

When the object is rotated using the goniometers, the vibration velocity vector $\mathbf{v_i}(t)$ rotates along with the object into a vector $\mathbf{v_i}(t)'$, where

$$\mathbf{v_i}(t)' = R(H_i, V_i) \cdot \mathbf{v_i}(t).$$

(B2)

$R(H, V)$ is the rotation matrix describing the rotation of the vector within the reference system of the interferometer. This rotation can be described as a first rotation about the y axis over an angle H, followed by a rotation over an angle V about the new position of the horizontal goniometer axis, resulting in

$$R(H, V) = \begin{vmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{vmatrix}$$

$$= \begin{vmatrix} \cos(H) & 0 & \sin(H) \\ \sin(H)\sin(V) & \cos(V) & -\cos(H)\sin(V) \\ -\sin(H)\cos(V) & \sin(V) & \cos(H)\cos(V) \end{vmatrix}$$

(B3)

[the matrix shown here is a corrected version for the one in Decraemer et al. (1994a), where Eq. (B3) was not correctly printed].

For observation point i, the goniometers were at angles H_i and V_i and the corresponding rotation matrix $R(H_i, V_i)$ will be denoted R^i.

With the interferometer we observe only the component of vibration along the interferometer's viewing direction that coincides with our z axis,

$$v'_{iz} = \begin{vmatrix} R^i_{31} & R^i_{32} & R^i_{33} \end{vmatrix} \cdot \begin{vmatrix} v_{tx} \\ v_{ty} \\ v_{tz} \end{vmatrix}$$

$$+ \begin{vmatrix} R^i_{31} & R^i_{32} & R^i_{33} \end{vmatrix} \cdot \begin{vmatrix} z_i\Omega_y - y_i\Omega_z \\ x_i\Omega_z - z_i\Omega_x \\ y_i\Omega_x - x_i\Omega_y \end{vmatrix}.$$

(B4)

For all observation points ($i = 1, n$) we can write the following matrix equation:

$$
\begin{vmatrix}
v'_{1z} \\
v'_{2z} \\
\cdots \\
v'_{iz} \\
\cdots \\
v'_{nz}
\end{vmatrix}
=
\begin{vmatrix}
R^1_{31} & R^1_{32} & R^1_{33} & (y_1 R^1_{33} - z_1 R^1_{32}) & (z_1 R^1_{31} - x_1 R^1_{33}) & (x_1 R^1_{32} - y_1 R^1_{31}) \\
R^2_{31} & R^2_{32} & R^2_{33} & (y_2 R^2_{33} - z_2 R^2_{32}) & (z_2 R^2_{31} - x_2 R^2_{33}) & (x_2 R^2_{32} - y_2 R^2_{31}) \\
\cdots \\
R^i_{31} & R^i_{32} & R^i_{33} & (y_i R^i_{33} - z_i R^i_{32}) & (z_i R^i_{31} - x_i R^i_{33}) & (x_i R^i_{32} - y_i R^i_{31}) \\
\cdots \\
R^n_{31} & R^n_{32} & R^n_{33} & (y_n R^n_{33} - z_n R^n_{32}) & (z_n R^n_{31} - x_n R^n_{33}) & (x_n R^n_{32} - y_n R^n_{31})
\end{vmatrix}
\cdot
\begin{vmatrix}
v_{tx} \\
v_{ty} \\
v_{tz} \\
\Omega_x \\
\Omega_y \\
\Omega_z
\end{vmatrix} .
\tag{B5}
$$

This is an equation with the measured velocities in the left-hand side matrix and the design matrix (function of the angles H_i, V_i and the coordinates x_i, y_i, z_i which we will denote by I) multiplied by the matrix with the six unknowns at the right-hand side.

In the experiments we use a harmonic sound stimulus and as a consequence all functions of time t, namely v'_{iz}, v_{tx}, v_{ty}, v_{tz}, Ω_x, Ω_y, and Ω_z, become also harmonic and we can write

$$v'_{iz} = A'_i \sin(\omega t + \vartheta_i),$$

$$v_{tx} = v^0_{tx} \sin(\omega t + \varphi_{tx}), \quad v_{ty} = v^0_{ty} \sin(\omega t + \varphi_{ty}),$$

$$v_{tz} = v^0_{tz} \sin(\omega t + \varphi_{tz}),$$

$$\Omega_x = \Omega^0_x \sin(\omega t + \xi_x), \quad \Omega_y = \Omega^0_y \sin(\omega t + \xi_y), \tag{B6}$$

$$\Omega_z = \Omega^0_z \sin(\omega t + \xi_z).$$

Substituting these expressions in Eq. (B5), expanding the sine functions with the sum formula and equating the $\sin \omega t$ and $\cos \omega t$ terms results in the following two sets of equations:

$$
\begin{vmatrix}
A'_{1z} \cos \vartheta_1 \\
A'_{2z} \cos \vartheta_2 \\
\cdots \\
A'_{iz} \cos \vartheta_i \\
\cdots \\
A'_{nz} \cos \vartheta_n
\end{vmatrix}
= I \cdot
\begin{vmatrix}
v^0_{tx} \cos \varphi_{tx} \\
v^0_{ty} \cos \varphi_{ty} \\
v^0_{tz} \cos \varphi_{tz} \\
\Omega^0_x \cos \xi_x \\
\Omega^0_y \cos \xi_y \\
\Omega^0_z \cos \xi_z
\end{vmatrix}
\tag{B7a}
$$

and

$$
\begin{vmatrix}
A'_{1z} \sin \vartheta_1 \\
A'_{2z} \sin \vartheta_2 \\
\cdots \\
A'_{iz} \sin \vartheta_i \\
\cdots \\
A'_{nz} \sin \vartheta_n
\end{vmatrix}
= I \cdot
\begin{vmatrix}
v^0_{tx} \sin \varphi_{tx} \\
v^0_{ty} \sin \varphi_{ty} \\
v^0_{tz} \sin \varphi_{tz} \\
\Omega^0_x \sin \xi_x \\
\Omega^0_y \sin \xi_y \\
\Omega^0_z \sin \xi_z
\end{vmatrix} .
\tag{B7b}
$$

Using five observation directions with six to eight observation points each, these equations represent overdetermined systems of linear equation. They are solved separately for $v^0_{tx} \cos \varphi_{tx}$, $v^0_{ty} \cos \varphi_{ty}$, $v^0_{tz} \cos \varphi_{tz}$, $\Omega^0_x \cos \xi_x$, $\Omega^0_y \cos \xi_y$, and $\Omega^0_z \cos \xi_z$ (these are the components of the "in-phase" translation and rotation vectors $\mathbf{v_{tc}}$ and $\mathbf{\Omega_c}$—subscript c)

and their sine counterparts $v^0_{tx} \sin \varphi_{tx}$, $v^0_{ty} \sin \varphi_{ty}$, $v^0_{tz} \sin \varphi_{tz}$, $\Omega^0_x \sin \xi_x$, $\Omega^0_y \sin \xi_y$, and $\Omega^0_z \sin \xi_z$ ("$\pi/2$ out-of-phase" vectors $\mathbf{v_{ts}}$ and $\mathbf{\Omega_s}$—subscript s). They are finally used to determine the six amplitudes ($v^0_{tx}\ v^0_{ty}\ v^0_{tz}\ \Omega^0_x\ \Omega^0_y\ \Omega^0_z$) and six phases ($\varphi_{tx}\ \varphi_{ty}\ \varphi_{tz}\ \xi_x\ \xi_y\ \xi_z$) that characterize the translation and rotation velocities. This procedure is repeated for each experimental frequency.

Let us note also that we can write $\mathbf{v}$ as $\mathbf{ds}/dt$ and $\mathbf{d\Omega}$ as $\mathbf{d\Theta}/\mathbf{dt}$ so that multiplying both sides of Eq. (B1) by dt we end up with an equation between infinitesimal linear and angular displacements

$$\mathbf{ds_i}(t) = \mathbf{ds_t}(t) + \mathbf{d\Theta}(t) \times \mathbf{r_i}. \tag{B1'}$$

The middle ear vibration amplitudes are so small (of the order of 10^{-8} m per Pa for translation and 10^{-5} rad per Pa for rotation) that Eq. (B1′) can be used to describe the experimental displacements when we let the differentials denote the very small but finite displacements. It follows that the equations derived here for the velocities also hold to determine the translational and rotational displacements, if so preferred.

APPENDIX C: TRANSFORMING THE MOTION COMPONENTS INTO THE INTRINSIC FRAME

To bring the stapes in an intrinsic reference frame it is first rotated and then a shift of origin is applied. When we describe the rotation by a matrix $R_{\mathrm{exp\_to\_int}}$, the in-phase and out-of-phase translation and rotation vectors $\mathbf{v_{tc}}$, $\mathbf{\Omega_c}$, and $\mathbf{v_{ts}}$, $\mathbf{\Omega_s}$ transform as

$$\mathbf{v_{tc\_rot}} = R_{\mathrm{exp\_to\_int}} \cdot \mathbf{v_{tc}}, \tag{C1a}$$

$$\mathbf{v_{ts\_rot}} = R_{\mathrm{exp\_to\_int}} \cdot \mathbf{v_{ts}}, \tag{C1b}$$

$$\mathbf{\Omega_{c\_rot}} = R_{\mathrm{exp\_to\_int}} \cdot \mathbf{\Omega_c}, \tag{C2a}$$

$$\mathbf{\Omega_{s\_rot}} = R_{\mathrm{exp\_to\_int}} \cdot \mathbf{\Omega_s}. \tag{C2b}$$

When the origin is shifted from O to O' (O' is chosen at the footplate center), a point i has a new position vector $\mathbf{r'_i}$, with

$$\mathbf{r'_i} = \mathbf{r_i} - \mathbf{OO'}. \tag{C3}$$

Substituting Eq. (C3) in Eq. (B1)

$$\mathbf{v_i} = \mathbf{v_t} + \mathbf{\Omega} \times (\mathbf{r'_i} + \mathbf{OO'}) = (\mathbf{v_t} + \mathbf{\Omega} \times \mathbf{OO'}) + \mathbf{\Omega} \times \mathbf{r'_i} \tag{C4}$$

or

$$\mathbf{v_i} = \mathbf{v_t'} + \mathbf{\Omega} \times \mathbf{r_i'} \tag{C5}$$

with

$$\mathbf{v_t'} = \mathbf{v_t} + \mathbf{\Omega} \times \mathbf{OO'}. \tag{C6}$$

Equation (C5) shows that the shift does not affect the rotation velocity, hence

$$\Omega_{c\_\text{int}} = \Omega_{c\_\text{rot}} \tag{C7a}$$

and

$$\Omega_{s\_\text{int}} = \Omega_{s\_\text{rot}}, \tag{C7b}$$

while Eq. (C6) shows how the translation velocity is transformed. Applying Eq. (C6) to Eqs. (C1a) and (C1b) yields finally

$$\mathbf{v_{tc\_int}} = \mathbf{v_{tc\_rot}} + \mathbf{\Omega_{c\_int}} \times \mathbf{OO'} \tag{C8a}$$

and

$$\mathbf{v_{ts\_int}} = \mathbf{v_{ts\_rot}} + \mathbf{\Omega_{c\_int}} \times \mathbf{OO'}. \tag{C8b}$$

Equations (C7) and (C8) can then be used to calculate the amplitude and phase of the translation and rotation velocities in the intrinsic frame of the stapes.

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.

Ball, G. R., Huber, A., and Goode, R. L. (**1997**). "Scanning laser Doppler vibrometry of the middle ear ossicles," Ear Nose Throat J. **76**, 213–222.

Besl, P. J., and McKay, N. D. (**1992**). "A method for registration of 3-D shapes," IEEE Trans. Pattern Anal. Mach. Intell. **14**, 239–256.

Decraemer, W. F., de La Rochefoucauld, O., Olson, E. S., Khanna, S. M., and Dirckx, J. J. J. (**2006**). "Estimation of stapes piston motion from uniaxial interferometer measurements along observation directions at an angle with the piston axis is prone to substantial errors," in *Vibration Measurements by Laser Techniques: Advances and Applications*, edited by E. P. Tomasini (SPIE, Bellingham, WA), Vol. **6345**, Paper No. 6345_12 (CD-ROM).

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. (**1996**). "Malleus vibrations in the cat ear are three dimensional," Proceedings of the Diversity in Auditory Mechanics, Berkeley, CA, pp. 115–121.

Decraemer, W. F., and Khanna, S. M. (**1997**). "Vibrations on the malleus measured through the ear canal," Proceedings of Middle Ear Mechanics in Research and Otosurgery, Dresden, 19–22 September **1996**, edited by K. B. Hüttenbrink, pp. 32–39.

Decraemer, W. F., and Khanna, S. M. (**1999**). *New Insights in the Functioning of the Middle-Ear. The Function and Mechanics of Normal, Diseased and Reconstructed Middle-Ears*, edited by J. J. Rosowski and S. Merchant (Kugler, The Hague, The Netherlands), pp. 23–38.

Decraemer, W. F., and Khanna, S. M. (**2000**). "Three-dimensional vibration of the ossicular chain in the cat," *Vibration Measurements by Laser Techniques: Advances and Applications*, edited by E. P. Tomasini (SPIE, Bellingham, WA), Vol. 4072, pp. 401–411.

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.

Decraemer, W. F., Khanna, S. M., and Funnell, W. R. J. (**1994a**). "A method for determining three-dimensional vibration in the ear," Hear. Res. **77**, 19–37.

Decraemer, W. F., Khanna, S. M., and Funnell, W. R. J. (**1994b**). "Bending of the manubrium in cat under normal sound stimulation," in *Optical and Imaging Techniques*, edited by H. J. Foth, A. Lewis, H. Podbielska, M. Robert-Nicoud, H. Schneckenburger, and A. Wilson, Proceedings of Progress in Biomedical Optics, Europto Series, 8–9 September **1994**, Lille, France (SPIE, Bellingham, WA), Vol. 2329, pp. 74–84.

Decraemer, W. F., Khanna, S. M., Funnell, and W. R. J. (**2000**). "Measurement and modelling of the three-dimensional vibrations of the stapes in cat," *Proceedings of the Symposium on Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 36–43.

Dong, W., and Olson, E. S. (**2006**). "Middle ear forward and reverse transmission in gerbil," J. Neurophysiol. **95**, 2951–2961.

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**, 140–152.

Heiland, K. E., Goode, R. L., Masanori, A., and Huber, A. (**1999**). "A human temporal bone study of stapes footplate movement," Am. J. Otol. **20**, 81–86.

Johnstone, J. R., Alder, V. A., Johnstone, B. M., Robertson, D., and Yates, G. K. (**1979**). "Cochlear action potential threshold and single unit thresholds," J. Acoust. Soc. Am. **65**, 254–257.

Khanna, S. M., Koester, C. J., Willemin, J. F., Dandliker, R., and Roskothen, H. (**1996**). "A noninvasive optical system for the study of the function of the inner ear in living animals," edited by V. Tuchin, *Selected Papers on Coherence Domain Methods in Biomedical Optics* (SPIE, Bellingham, WA), Vol. 2732, pp. 64–81.

Khanna, S. M., and Stinson, M. R. (**1985**). "Specification of the acoustical input to the ear at high frequencies," J. Acoust. Soc. Am. **72**, 577–589.

Kurokawa, H., and Goode, R. L. (**1995**). "Sound pressure gain produced by the human middle ear," Otolaryngol.-Head Neck Surg. **113**, 349–355.

Lynch, T. J., Nedzelnitsky, V., and Peake, W. T. (**1982**). "Input impedance of the cochlea in cat," J. Acoust. Soc. Am. **72**, 108–130.

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.

Olson, E. S., and Cooper, N. (**2000**). "Stapes motion and scala vestibuli pressure," Abstracts of the 23rd Annual Midwinter Research Meeting, St-Petersburg Beach, FL, Abstract No. 399. Accessible at the ARO website.

Overstreet, E. H. III, and Ruggero, M. A. (**2002**). "Development of wide-band middle ear transmission in the Mongolian gerbil," J. Acoust. Soc. Am. **111**, 261–270.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1989**). *Numerical Recipes* (Cambridge University Press, New York).

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**, 1–17.

Ravicz, M. E., and Rosowski, J. J. (**2004**). "High-frequency sound transmission throughthe gerbil middle ear," Abstracts of the 27th Annual Midwinter Research Meeting, Abstract No. 817. Accessible at the ARO website.

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.

Rosowski, J. J., Ravicz, M. E., Teoh, S. W., and Flandermeyer, D. (**1999**). "Measurement of middle-ear function in the Mongolian gerbil, a specialized mammalian ear," Audiol. Neuro-Otol. **4**, 129–136.

Teoh, S. W., Flandermeyer, D. T., and Rosowski, J. J. (**1997**). "Effects of pars flaccida on sound conduction in ears of Mongolian gerbil: acoustic and anatomical measurements," Hear. Res. **106**, 39–65.

Willemin, J. F., Dandliker, R., and Khanna, S. M. (**1988**). "Heterodyne interferometer for submicroscopic vibration measurements in the inner ear," J. Acoust. Soc. Am. **83**, 787–795

Basilar membrane mechanics in the 6–9 kHz region of sensitive chinchilla cochleae

William S. Rhode[a]

Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706

The vibration of the basilar membrane in the 6–9 kHz region in the chinchilla cochlea has been studied using a displacement sensitive interferometer. Displacements of 0.7–1.4 nm at 0 dB sound pressure level have been obtained. At the characteristic frequency (CF), rate-of-growth (ROG) functions computed as the slope of input-output (IO) functions can be as low as 0.1 dB/dB. IO functions for frequencies >CF have ROGs near 0 dB/dB and can have notches characterized by both negative slopes and expansive ROGs, i.e., >1 dB/dB. For frequencies <0.6*CF, ROGs >1.2 dB/dB were found. Cochlear gain is shown to be greater than 60 dB in sensitive preparations with a single cochlea having nearly 80 dB gain. The compressive nature of the cochlea remains at all levels though it is masked at frequencies >CF when the amplitude of a compression wave exceeds that of the traveling wave. The compression wave produces the plateau region of the mechanical response at high intensities and has a nearly constant phase versus frequency function implying a high velocity. The summation of the traveling and compression waves explains the occurrence of the notches in both the IO and iso-intensity functions. Vibration of the osseous spiral limbus may alter the drive to inner hair cells. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2718397]

I. INTRODUCTION

Cochlear mechanical vibration has been studied in the high frequency region of the cochlea in several species (reviewed in Robles and Ruggero, 2001). Because chinchillas have a similar frequency response range to that of humans they have been frequently used for hearing related studies. Coiling of the mammalian cochlea and the fact that it is partially if not largely buried in the temporal bone, makes if difficult to access all but the high frequency region, the initial few mm. Measurements made in the 9–20 kHz region of the chinchilla cochlea demonstrate that the tuning and sensitivity are adequate to explain most of the same auditory nerve features (Robles *et al.*, 1986). There may remain some small differences as shown by Narayan *et al.* (1998) in the only study that performed auditory nerve (AN) and basilar membrane (BM) mechanics in the same preparation. There remains the problem of accessing lower frequency regions that overlap with the speech-related frequency range (0.1–5 kHz) though there have been a few studies of the very apical (low frequency) cochlear region (e.g., Rhode and Cooper, 1996; Khanna and Hao, 1999; Morioka *et al.*, 1995). The goal of this study was to access the BM region having the lowest possible characteristic frequency (CF) using the open bulla approach to the cochlea and to characterize the basilar membrane vibration in very sensitive preparations. While CFs as low as 5 kHz have been recorded from that region, ~3.5–4.25 mm from the base in this study, the most sensitive cochleae had CFs in the range of 6–9 kHz. Note that relying exclusively on the distance from the base as a

measure of the CF results in a flaw in estimating CF in that the length of the chinchilla cochlea varies from 16 to 25 mm in different animals. Note that cochlear features and tonotopicity are constant between cochleae when compared as a percentage of length (Carr and Bohne, 1979).

The nonlinear, compressive behavior of BM vibration has been extensively documented (Rhode, 1971; Sellick *et al.*, 1982; Robles *et al.*, 1986). The nonlinearity, a region of compressive amplitude growth on the order of 0.3 dB/dB at CF, has been described as limited to less than an octave range centered around CF. The present study demonstrates that BM vibration is nonlinear at all frequencies with an expansive nonlinearity at frequencies not in the compressive frequency region.

Previous studies have addressed distortion or sometimes the lack of distortion in cochlear vibrations. The largest distortion magnitudes occur at the CF recording site for the 2nd harmonic of a stimulus frequency that is equal to one-half of CF (Cooper, 1998). Because cochlear distortion has to be a function of the magnitude of cochlear nonlinearity, it is essential to make distortion measurements in sensitive cochleae.

Only a few studies have been reported addressing the vibration of the osseous spiral limbus or OSL (e.g., Rhode, 1971). It was reported that the BM vibrated at least 20 dB more than the OSL over the entire frequency range though OSL vibration measurements were not made for frequencies greater than CF in what is called the plateau region. If the OSL vibrated nearly as much as the BM at frequencies in the "plateau" region, excitation of the inner hair cells would be reduced thereby accounting for the lack of a plateau in auditory nerve (AN) fiber recordings. This is of interest because no explanation of the existence of the "plateau" region in the

[a] Electronic mail: rhode@physiology.wisc.edu

BM responses above CF has been presented other than the hypothesis that it could be due to a fast compressive wave (Rhode, 1971; Peterson and Bogert, 1950). Notable is the occurrence of a pressure and phase plateau recorded near the BM with a very small pressure sensor that corresponds to the vibration plateau (Olson, 1998). Recently, evanescent waves have been introduced into the discussion of cochlear mechanics as they also have constant phase as a function of frequency implying they have high velocity (see Discussion Session, Nuttall *et al.*, 2006).

II. METHODS

Methods are essentially those detailed in Cooper and Rhode (1992). Data from 48 chinchilla cochleae were studied at 3.5–4.25 mm from the basal end of the basilar membrane with CFs in the 5–9 kHz range. Twenty-seven cochleae had a BM displacement sensitivity expressed as "dB re 1 nm at 1 Pa" >78 dB and seven had a sensitivity >90 dB with one >96 dB. All procedures were approved by the Animal Care and Use Committee of the University of Wisconsin.

Each animal was anesthetized with pentobarbital using a dose rate of 75 mg/kg. Additional doses were administered to maintain the animal in a deeply areflexive state. All anesthetics were administered intraperitoneally. A tracheotomy was performed to ensure an open airway and to place the animal on a respirator if necessary, though it was never used. After the pinna was surgically removed, four screws were implanted in the skull and cemented in with dental cement in order to form a rigid base. A bolt was then cemented to the dental cement base to provide a stable fixation of the skull to a head holder with six degrees of freedom for the purpose of positioning the cochlea under the microscope.

In early experiments, the bulla was opened widely and a silver ball electrode was positioned so as to touch the edge of the round window for the purpose of recording the compound action potential (CAP) of the auditory nerve in response to short duration tones (16 ms) for each animal. The stimulus frequency was stepped in 2 kHz increments from 2 to 20 kHz. At each frequency a visual detection threshold for CAP was determined by viewing an average of 20 repetitions as the stimulus level was varied in 1 dB steps. If the thresholds were above our best threshold curve by more than 30 dB, no data were collected, as high CAP thresholds equated to little or no compression in the hook region (Sellick *et al.*, 1982). In later experiments no CAPs were recorded and no middle-ear muscles were cut (in any experiments) in order to avoid any possible traumatic effects of preparing the cochlea. The amount of cochlear nonlinearity at CF was used as the exclusive monitor of the state of the preparation. The sensitivity of the preparation at CF was measured by repeated input-output (IO) functions and in a few instances was maintained over a 12 h period.

The overlying cochlear bone in a region with CFs between 5 and 9 kHz was shaved down using a microdrill (0.45 mm burr) until the remaining tissue and/or bone debris could be removed with a pick fabricated out of a tungsten microelectrode or flushed out with a stream of artificial peri-

lymph. Gold-coated polystyrene beads 25 μm in diameter served as retroreflectors. They were placed in the perilymph and allowed to sink to the basilar membrane, and as such their final resting place was uncontrolled. They have a specific gravity of 1.05 that is near that of water (1.0) and therefore any loading of the basilar membrane by the bead should be minimal as shown by Cooper (2000). A glass cover slip was placed over the cochlear opening with no hydromechanical seal. The cover glass served to avoid the problem of an unstable air-fluid interface which can alter the measurements (Cooper and Rhode, 1992). Occasionally, several beads were successfully placed along the BM, which allowed estimates of traveling wave velocity. When beads fell on the OSL near (50–100 μm) the basilar membrane, its motion was also measured (22 cases).

A displacement-sensitive interferometer was used to measure the vibration (Cooper and Rhode, 1992). Measurements <0.01 nm could be made at the higher frequencies (>5 kHz) while $1/f$ noise becomes an increasing problem as the frequency decreases. There were no bandwidth limitations for the frequency range investigated here. The signals were sampled at 250 kHz and averaged wave forms were saved. All analysis routines were MATLAB® based. A software phase-locked analysis of the displacement wave form was performed after application of a Hamming data window. An exact number of periods of the stimulus frequency were included in the analyzed segment.

An opening in the bony ear canal, immediately over the tympanic membrane, was made so that either an Etymotic phone/microphone or a Shure 5 phone that was acoustically calibrated using a probe tube to which a $\frac{1}{2}$ in. Bruel & Kjaer condenser microphone is adjoined. The probe tube was visualized as it was positioned above the tympanic membrane within 1–2 mm of the tip of the malleus. The opening was sealed with a glass cover after a 45 μm bead was placed on the tympanic membrane at the tip of the malleus (umbo). The bead was used as a retroreflector for the interferometer and allowed the measurement of the transfer function of the malleus. Stimuli were presented using a TDT System I and purpose-built software.

Stimuli consisted of 4–10, 30 ms tone pips presented every 105 ms. Intensity usually covered 100–0 dB sound pressure level (SPL) in −5 dB steps though the maximum intensity was limited by the range of the analog-to-digital (A/D) converter. Frequency steps were typically 200 Hz up to the CF region, where they were changed to 100 Hz for all higher frequencies. Click stimuli were produced by a $\frac{1}{2}''$ condenser microphone and 1000 reps were averaged. Peak SPL (pSPL) for clicks was defined as the equivalent SPL at 1 kHz based on the maximum voltage measured with the microphone for the click. At maximum output the pSPL was 109 dB.

III. RESULTS

A. Input/output (IO) and iso-intensity functions

Five of sixty IO functions for a study in which the CF was 6.6 kHz illustrate the range of IO functions for typical BM vibration measurements [Fig. 1(a)]. For frequencies
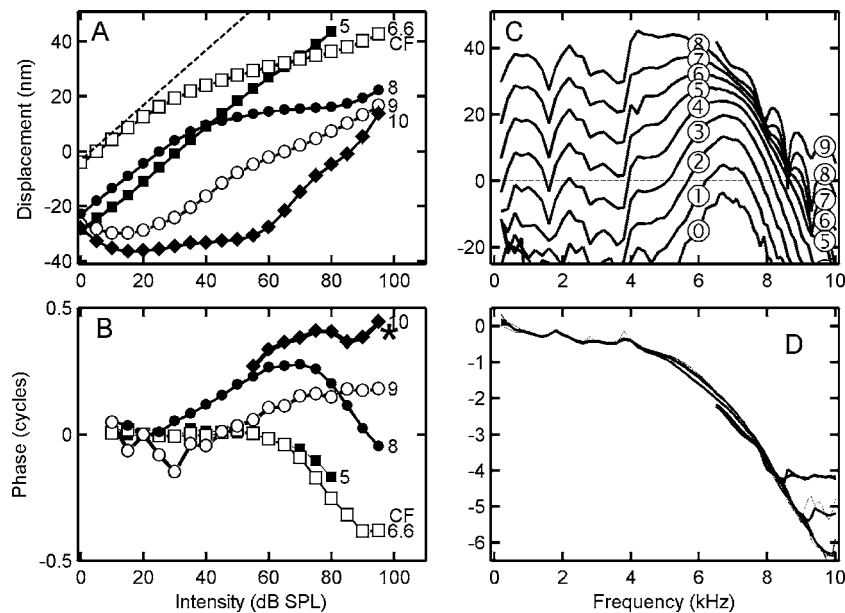
FIG. 1. IO and transfer functions for a BM location with a CF=6.6 kHz. A. IO functions for a frequency below CF (5 kHz), CF (6.6 kHz), and three frequencies above CF. The dashed line corresponds to linear growth of vibration and indicates that at CF there is a deviation from linearity beginning near 0 dB SPL. B. Corresponding IO phase curves for the IO displacement functions in A normalized by their values at 20 dB SPL with the 10 kHz curve shifted up for clarity. C. Frequency transfer functions (TFs) constructed from the IO functions shown at multiples of 10 dB SPL as indicated by the symbols (e.g., ③=30 dB SPL). The dips at ~1.5 and 3.5 kHz are from the middle-ear transfer function under the open-bulla condition. D. the phase TFs corresponding to the displacement TFs in C. Data were collected at 68 frequencies and 20 intensities, 0–95 dB SPL in 5 dB steps. Measurements were made at six locations with CFs ranging from 6 to 6.8 kHz over a 12 h period with nearly identical results except for the CF. The arrow indicates where CF is located and where there is no change in phase below 60 dB SPL. The thickest line corresponds to the phase at 90 dB SPL. Chinchilla N92.

more than 0.5 octaves below CF the IO functions are all linear. As the stimulus frequency approaches CF, e.g., 5 kHz, the IO functions become nonlinear at high levels and are accompanied by a phase lag [Fig. 1(b)]. CF is defined as the frequency for which the ordinate axis is intersected at the highest value of all IO intersections. In this case CF is 6.6 kHz and the intersection is at 0.7 nm at 0 dB SPL. The deviance of the IO function at CF from the dashed line (linear IO) indicates nonlinearity at and likely below 0 dB SPL. This IO function has two main slopes with a transition between the two in the vicinity of 20–30 dB SPL. There is no indication of the nonlinearity changing significantly below 100 dB SPL. This is important because often there is a return to linearity in the IO functions at high levels, and it is often concluded on theoretical grounds that it should (Goldstein, 1995). For frequencies just above CF, IO functions are nonlinear at all intensities and have ranges of intensities where the rate of growth (ROG) of the displacement is near 0 dB/dB (e.g., 8 kHz). At higher frequencies, a region is reached where the IO functions become linear at high intensities (e.g., 10 kHz). This has been called the "plateau region" of the response (Rhode, 1971) and it does not appear to be the result of any instrumental artifact (Robles and Ruggero, 2001). However, note that the IO functions in this region are compressive at low levels.

Frequency transfer functions (TFs) combine all 60 IO functions into a displacement-vs-frequency display though only every other intensity is illustrated for clarity in Fig. 1(c). BM-TFs include the middle-ear response, whose dips account for the dips in the BM functions at 1.5 and 3.5 kHz (the middle-ear response is not removed because the hair cells are driven by the combined effects of the middle ear and the BM). At 0 dB SPL, the peak is at 6.6 kHz and the width of the curve 10 dB down from the peak yields Q_{10} =CF/bandwidth (10 dB above threshold)=6600/1488=4.4. This is the quality factor for this filter and is comparable to that of auditory nerve filters with like CFs which typically have Q_{10}'s in the range of 3.3–6.5 (e.g., Goldstein, 1995). For frequencies low relative to CF, the TFs increase linearly as a function of intensity. Near CF the intensity TFs are closer together due to the compressive nonlinearity and at certain frequencies above CF have characteristic dips or notches. For frequencies high relative to CF, there is a region where the TFs increase at a linear rate with increasing intensity.

Each IO [Fig. 1(a)] and TF displacement function [Fig. 1(c)] has a corresponding phase characteristic shown in Figs. 1(b) and 1(d), respectively. At low frequencies there is negligible phase change with increasing intensity. As the CF of 6.6 kHz is approached and compression begins, there is a phase lag with increasing intensity [e.g., 5 kHz, Fig. 1(b)]. At CF there is little phase change at low intensities and a phase lag that approaches 180° at high intensities. For frequencies above CF, there is an initial phase lead that rarely exceeds 90° with increasing intensity that is followed by a phase lag that can reach 180°. The phase lag/lead below/above CF was originally reported by Anderson et al. (1971) in squirrel monkey AN recordings. The principal difference between the present results and theirs is that for all frequencies in the nonlinear region there is a phase lag at high intensities in the present data. There is a corresponding TF phase characteristic that illustrates the same behavior [Fig.

William S. Rhode: Nonlinear mechanics of sensitive chinchilla cochleae
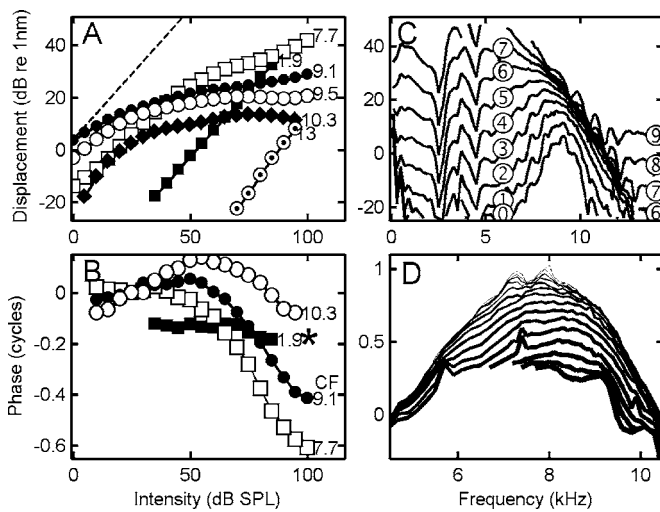
FIG. 2. IO and transfer functions for a location with a CF=9.1 kHz. Data were collected at 105 frequencies and 21 intensities, 0–100 dB SPL in 5 dB steps. Data were not always available at the high intensities when the signal amplitude would saturate the A/D converter. A. IO functions for two frequencies below CF (0.7 and 7.7 kHz), CF (9.1 kHz), and three frequencies above CF. Dashed lines correspond to linear growth of vibration and indicate that there is a deviation for CF responses from linearity even at 0 dB SPL and for both low and high frequencies relative to CF. B. Corresponding IO phase curves for the IO displacement functions in panel A. C. Frequency transfer functions (TFs) constructed from the IO functions shown at multiples of 10 dB SPL as indicated by the symbols (e.g., ③=30 dB SPL). The dips at ∼2.5 and 4.5 kHz are from the middle-ear transfer function. D. The phase TFs corresponding to the displacement TFs in C. In order to emphasize the phase behavior in the region of CF, the frequency range was truncated and the values of the phase-transfer function were set to 0° at 4.5 and 10.5 kHz. The latter two frequencies produced responses whose phases did not change significantly with intensity. Line thickness codes SPL with the thickest=90 dB SPL. N157.

1(d)]. At low frequencies all phase curves overlay one another. In the region of CF there is a spread of the curves that is characterized by little change at CF and low levels (indicated by the arrow) with all curves showing increased lag at high levels. For frequencies in the plateau region the phase attains a relatively constant value that varies with level in increments of one cycle. The implication of this will be discussed later.

For comparison with BM mechanics at a somewhat higher CF, the data for a 9.1 kHz location, in a cochlea that had the highest sensitivity (96 dB) in this study, is shown in Fig. 2. The IO functions follow a similar pattern as those in Fig. 1(a). At CF there is an ROG of ∼0.5 dB/dB at 0 dB SPL with a displacement of 1.4 nm [Fig. 2(a)]. Above 40 dB SPL, the ROG is ∼0.1 dB/dB. In addition, at 100 dB SPL, the IO function remains highly compressive though there is a small increase in the slope which is frequently seen in other experiments. One aspect of IO functions that stands out at both 0.7 and 13 kHz is that they are both expansive, ∼1.25 dB/dB. Expansive IO functions were present in 16 sensitive cochlea to various degrees and often limited to frequencies <1 kHz. This example indicates nonlinear behavior can extend over all frequencies that a cochlear region responds to. This point will be taken up again in the next section.

TFs in Fig. 2(c) follow the same pattern as those of Fig. 1(c), what "appears" to be linear steps with increasing inten-

sities at low frequencies, compressive steps around CF and a return to near linear steps at high frequencies and intensities. Also there is a ripple pattern in the functions for frequencies >CF and sufficiently high intensities. Q_{10}=6.3 for this location is about what is obtained for low-threshold AN fibers with the same CF though there is considerable variation in AN Q_{10}'s that largely correlates with the firing threshold (unpublished study of chinchilla AN tuning curves).

The IO phase curve at 7.7 kHz in Fig. 2(b) illustrates that the phase change with intensity can be greater than 220° (as much as a 360° change has been observed in another experiment). For a phase-versus-frequency display the phase at 4.5 and 10.5 kHz was set to 0° in order to emphasize the phase-versus-intensity behavior in Fig. 2(d). There is ∼250° phase change with intensity at 8 kHz, ∼170° at CF and ∼90° at 10.3 kHz. For either higher or lower frequencies than the range shown, the phase is constant as a function of intensity. For frequencies beyond CF, there are phase leads at low levels and phase lags at high levels [Fig. 2(b), 10.3 kHz].

B. Sensitivity and rate-of-growth (ROG) functions

Cochlear mechanical sensitivity functions are defined here as the displacement at each intensity extrapolated to 1 Pas (94 dB SPL) using linear extrapolation. That is, response amplitude at a dB SPL specified by "X" for each frequency is normalized by X (i.e., (94-X)dB is added). Note that in a linear system, all the iso-intensity functions would superimpose.

Sensitivity functions for the two sets of data in Figs. 1(c) and 2(c) are shown in Figs. 3(a) and 4(a), respectively. In the regions of linear vibration all the curves overlay one another. In the region of nonlinearity there is increasing separation from the highest curve obtained at 0 dB SPL, with each 5 dB increase in intensity. In each case, the highest sensitivity of ∼90 dB re 1 nm occurs at the lowest intensity, 0 dB SPL, and would result in a displacement of >60 μm at 100 dB SPL if there were no compression of the displacements [Fig. 3(a)]. Twenty-eight studies had >80 dB sensitivity at CF, of which six surpassed 90 dB in sensitivity. One study, shown in Fig. 4, had a sensitivity of 96 dB. Sensitivity is affected by the radial location of the reflective bead that has to be near the center of the BM for a maximal response. There is little control on achieving the optimum location as the beads freely sink onto the basilar membrane so that the maximal displacement is likely underestimated in most cases. An alternative representation of the sensitivity curves is when they are corrected for the middle-ear transfer function [Figs. 3(c) and 4(c)]. The gains at CF are 62 and 70 dB respectively. The slope of the low frequency portion of the TFs is usually between 6 and 7 dB/oct which is what would be expected for a first order filter.

Another characterization of BM vibration is rate-of-growth (ROG) of the displacement IO functions [Figs. 3(b) and 4(b)]. ROG is measured in dB/dB and is the slope of the IO functions in Figs. 1(a) and 2(a). For a linear system the ROG=1 dB/dB and for complete compression it=0 dB/dB. These two values are indicated by the horizontal dashed lines
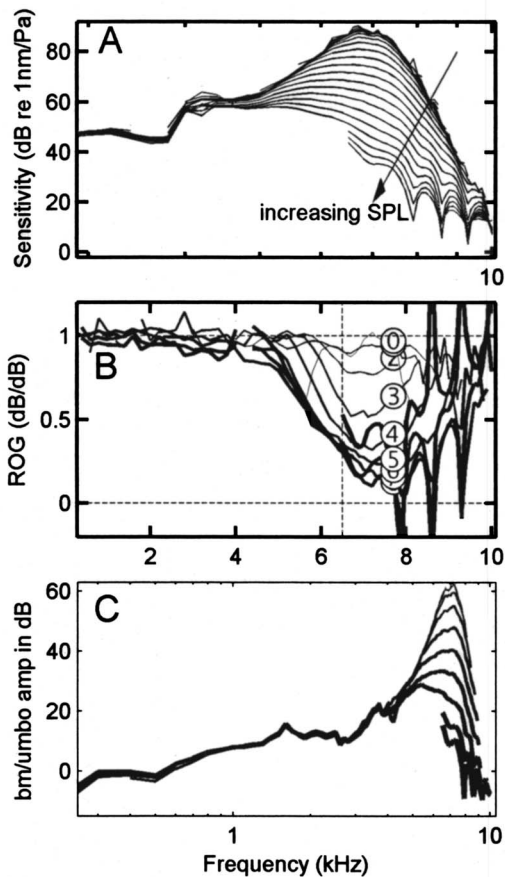
FIG. 3. A. Sensitivity functions for the data in Fig. 1(c). Noisy data at low amplitudes have been eliminated. Higher intensity curves are the lower ones. Intensity in 5 dB increments. B. Rate-of-growth (ROG) functions obtained from the IO functions in Fig. 1(a). The horizontal dashed lines indicate linearity (ROG=1) and complete compression (ROG=0). The vertical dashed line designates the CF. The low intensity (<60 dB) curves were three point smoothed (using filtfilt in MATLAB). Frequency was sampled every 100 Hz. Intensity was sampled every 5 dB though only every other curve is shown for clarity. C. Sensitivity functions relative to the middle-ear displacement.

in Figs. 3(b) and 4(b). The vertical dashed line is positioned at CF. Data are presented on a linear frequency scale to expand the curves in the CF region. In Fig. 3(b) at CF the ROG is ~0.9 dB/dB at threshold and decreases with increasing intensity to ~0.25 dB/dB. In an exceptionally sensitive and compressive cochlea with a CF=9.1 kHz the ROG is ~0.45 dB/dB at threshold and decreases with increasing intensity to ~0.08 dB/dB [Fig. 4(b)].

There are two frequency regions where the ROG is always described as linear, frequencies low relative to CF and higher than CF where the plateau region is. Below 4 kHz the ROG is near 1 dB/dB (linear) for all frequencies in Fig. 3(b) while in contrast there is an expansive nonlinearity that is as large as 1.25 dB/dB in this region in Fig. 4(b). The difference in the two studies was their cochlear sensitivities, 84–90 and 95 dB re 1 nm at 1 Pa, respectively, and ROG at CF ~0.1 *versus* 0.3 dB/dB, i.e., much greater compression in the second example. While the expansive nonlinearity is present for frequencies below 5 kHz [Fig. 4(b)], it is not uniform throughout what is often referred to as the tail region of the response. At frequencies higher than CF and suf-
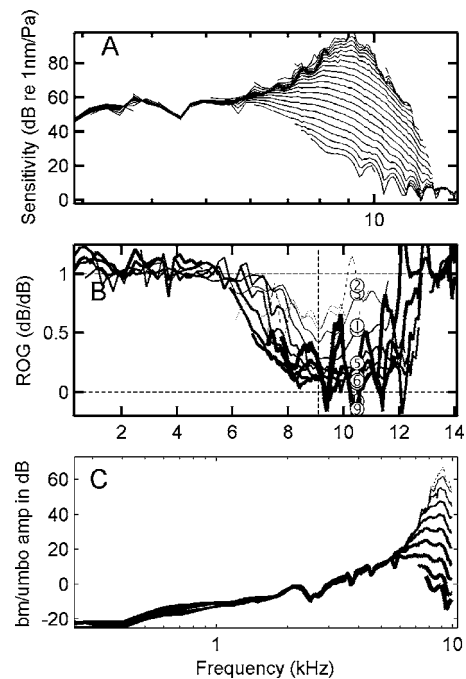
FIG. 4. A. The sensitivity functions for the data in Fig. 2(c). Noisy data have been suppressed. B. ROG functions obtained from the IO functions in Fig. 2(a). The horizontal dashed lines indicate linearity (ROG=1) and complete compression (ROG=0). The vertical dashed line designates the CF. The low intensity (<60 dB) curves were three point smoothed (filtfilt in MATLAB). Frequency was sampled every 100 Hz. Intensity was sampled every 5 dB though only every fourth curve is shown for clarity. C. Sensitivity functions relative to the middle-ear displacement.

ficiently high intensities, i.e., the plateau region, the ROG is always reported to be linear [Fig. 3(b)]. However, in the plateau region around 14 kHz, the ROG is 1.15–1.25 dB/dB which is in the same range of ROGs as found for frequencies in the tail region of the response [Fig. 4(b)].

At the frequency of the notches, a pattern is followed in which there is first a negative slope, then a minimum, followed by a rapid increase to >1 dB/dB, that is, an expansive ROG that can attain values of >1.5 dB/dB. The occurrence of the notches results in the ripple pattern in the displacement-frequency functions [Fig. 3(b)]. This ripple pattern is present even below CF in Fig. 4(b) at high intensities (thicker lines). Whether the source of the ripple pattern has any relation to the occurrence of ROG=1 dB/dB at ~2 and 4 kHz is an open question.

C. Notch frequencies

Notches in AN-fiber IO functions have been explained as resulting from two components driving the AN firing (Kiang *et al.*, 1986). There was a corresponding 180° phase shift above the notch intensity. Direct pressure measurements in the scala tympani near the BM obtained notches followed by a phase plateau that is interpreted as the interaction of a fast wave with the slower traveling wave (TW) (Olson, 1998). Notches can easily be missed if the sampling parameters are too crude to characterize them. That is, for CF ~6 kHz, a maximum of 5 dB amplitude steps and 100 Hz frequency steps (or pure luck) are required for adequate characterization though it is very likely that the exact notch frequency
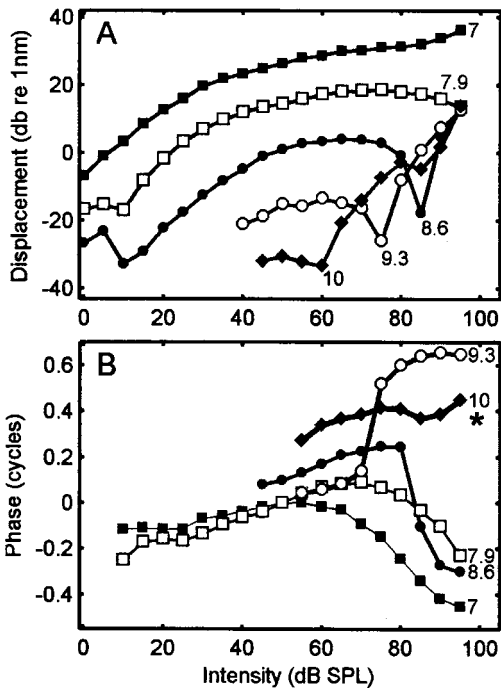
FIG. 5. A/B. amplitude and B. phase IO functions obtained from the animal of Fig. 1 near notch frequencies. The 10 kHz curve was shifted vertically for clearer display.

FIG. 6. Behavior of a notch when the frequency was stepped in 1 Hz increments and the intensity in 1 dB steps. Amplitude (solid lines) and phase (dashed lines) are shown for stimuli at 7897 and 7897 Hz. The deeper dip corresponds to 7897 Hz. The phase flipped at 7896 Hz (not shown). The phase change between 7895 and 7897 Hz is almost 270°.

will be missed in most instances. IO functions at the frequency of the notches shown in Fig. 1(c) are illustrated in Fig. 5(a). The most prominent notch occurs at 8.6 kHz where there is a 20 dB dip that is accompanied by over a 180° phase shift within a few dB increase in intensity [Fig. 5(b)]. At successively higher notch frequencies, the notch occurs at lower and lower intensities. In each case, a rapid phase shift of ∼180° with increasing intensity at the notch frequency occurs [Fig. 5(b)]. At 7 kHz the dip is very small [note the inflection at 7 kHz and 90 dB SPL in Fig. 1(c)] and is included to emphasize that there is a relatively systematic frequency separation of ∼700 Hz of the notches that appears to be determined by the slope of the phase-vs-intensity functions. That is, each time the BM-TW phase changes by a cycle it appears that a second BM vibration, the fast (compression) wave, adds in a destructive manner when its amplitude becomes just equal to the TW's amplitude. This proposed mechanism is a result of the flat phase curves at high frequencies where the "plateau" phases differ in one cycle amounts in Fig. 1(d). That is, because the fast-wave phase is nearly constant as a function of frequency (indicating high velocity), the rapid changing TW phase has to be opposite in phase at some frequency along with an amplitude match to result in a notch. It is worth repeating that the IO functions at notch frequencies have an expansive nonlinearity, i.e., growth rate >1 dB/dB for levels above the notch. This implies there is a nonlinear impedance in the OSL-cochlear partition complex at these high frequencies.

An example of a rapid amplitude (solid line) and phase transition (dotted line) for an IO function at a notch frequency is shown in Fig. 6 where the frequency step size was 2 Hz and the intensity step was 1 dB. In fact, increasing the

frequency one Hz caused the phase function to flip (not shown). The notch is over 40 dB deep indicating nearly complete cancellation of vibration.

D. Cochlear gain

The method of determining the cochlear gain or equivalently the compression of the system is somewhat an open question. This results from the fact that the IO function at CF usually does not become linear below 100 dB SPL. In the examples shown in Fig. 7, the intensity was increased to 100 dB SPL and yet the IO function did not become linear. The dot-dashed lines indicate linearity, the length of the horizontal dashed line is a measure of cochlear gain and indicates a value greater than 60 dB in Fig. 7(a) and about 78 dB in Fig. 7(b). Values greater than 60 dB were found in several instances and are likely present in additional instances where the intensity was not increased to sufficient levels to document the actual gain. That is, in many experiments the maximum intensity was limited to 80–90 dB SPL resulting in an underestimate of cochlear gain. In a few experiments where the intensity was increased even to 115 dB SPL, linearity of response was not achieved at CF.

An alternate method of estimating the cochlear gain is to measure CF-IO functions before and after euthanizing the animal. IO functions become "nearly linear" within a few minutes after cardiac arrest as shown in Fig. 7(c). A correlate of the rapid decrease in gain is the rapid loss of the endocochlear potential after administration of furosemide which appears to shut down cochlear amplification (Sewell, 1984).

Gain is estimated by the horizontal shift in the IO functions that occurs upon death. This technique has been previously used to estimate the gain by Nuttall and Dolan (1996) who reported a gain of as much as 65 dB in the high frequency region of a guinea pig cochlea. The IO function is described as nearly linear immediately after death because at CF it remains slightly nonlinear above 70 dB SPL for a variable period up to 1 h with the ROG= ∼0.9 dB/dB. This is an additional confounding feature in estimating the gain. However, the important point is that some aspect of the co-
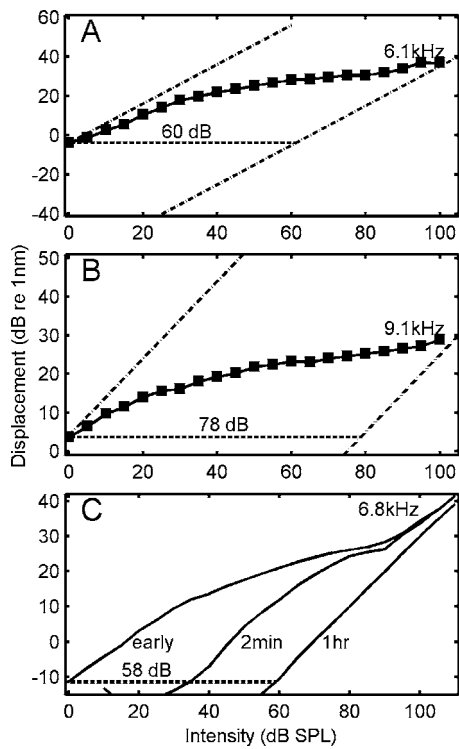
FIG. 7. IO functions at CF for three cochleae for estimating gain. A and B. Intensity was varied from 0 to 100 dB SPL in 5 dB steps. The dot-dashed lines indicate linear growth. Where the linear line on the right should be drawn is problematic because there is no indication of a return to linearity at high intensities. Gain is >60 dB in A (N136), approaches 80 dB in B (N157). C. IO functions for a live/dead comparison (N140). Gain estimated to be 58 dB; 2 min and 1 h are the times of the measurements after the animal's breathing stopped due to an overdose of anesthetic. In this instance the IO function approached linearity at high levels though the ROG was 0.9 dB/dB at high levels for over 30 min after death.

chlea remains viable and results in this much reduced cochlear nonlinearity, perhaps either tip link motors or electromotility due to prestin remains partially active. Numerous *in vitro* studies indicate that the outer hair cells (OHCs) may be viable for up to couple of hours after isolation.

E. Iso-displacement functions

The intersection of the respective response curves with the dashed horizontal line at 1 nm (0 dB) in Fig. 1(c) results in a 1 nm iso-displacement curve (=a mechanical threshold curve) that can be compared to frequency-threshold curves (FTCs) of AN fibers. If it is assumed that displacement is the adequate stimulus for transduction in the inner hair cells, one can compare the most sensitive AN-FTCs to the present mechanical iso-displacement curves. Two AN-FTCs are compared to the 6.5 kHz mechanical data in Fig. 8(a). They have a 10 and a 33 dB threshold, respectively, and there is close agreement with the 1 and 10 nm mechanical curves especially in the tip region, respectively. For frequencies below the compressive region for the BM mechanics (i.e., below 4 kHz) there is a significant deviation between the two functions that requires explanation. Comparing the BM iso-displacement curves and FTCs in Fig. 8(b) for the 8.1 kHz data, one could conclude that very little further processing beyond the BM is necessary to account for the FTCs. A BM
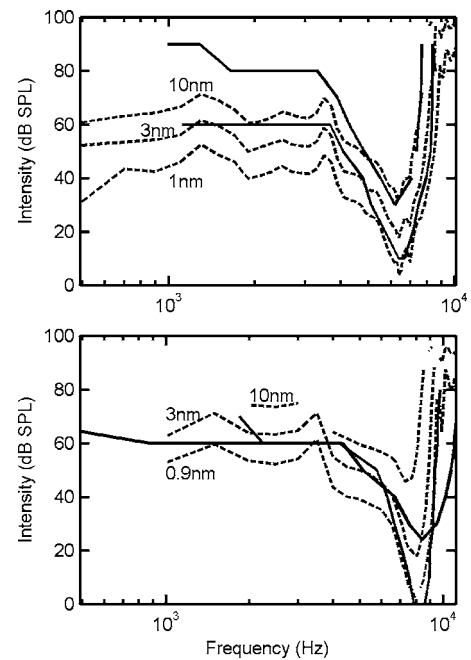
FIG. 8. A. Comparison of three iso-displacement curves (dashed lines) and two AN-fiber iso-rate curves (solid lines) obtained in a separate unpublished study for fibers with similar CFs to the mechanical data. CF= ∼6.1 kHz. N136. B. Comparison of three iso-displacement curves and two AN-fiber iso-rate curves with similar CFs to the mechanical data obtained in a separate study. Mechanical CF= ∼9.1 kHz. N79.

Q_{10}=6.3 is comparable to those found for in cat AN (Q_{10} between 4 and 6, Rhode and Smith, 1985). However, in general, there is a discrepancy between the tip-tail difference in FTCs and the mechanical tip-tail values, being smaller in the latter by about 10 dB. This may relate to the difference in the two preparations, i.e., an open cochlea and bulla for mechanical measurements versus a closed cochlea and bulla for the AN studies or merely the condition of the preparations. Narayan *et al.* (1998) noted similar differences between ANF and BM-tuning curves when both measurements were performed in the same preparation. However, they found that neither the displacement or velocity BM tuning curves "exactly" fit the AN-FTCs.

F. Harmonic distortion

One prominent feature of signal processing of signals in the cochlea is the compressive nonlinearity. Considering the degree of compression, the magnitude of harmonic distortion appears to be modest. Prominent distortion products are the second and third harmonics of the fundamental frequency. Maximum distortion is generated when the frequency of the stimulus is a half or a third the CF of the cochlear location under study (i.e., f=CF/n where n=2 or 3). For example, in Fig. 9(a), with CF= ∼6 kHz the second harmonic is typically between −15 and −30 dB re the amplitude of the fundamental while the third harmonic is typically about 5−10 dB less than the second harmonic. In absolute magnitude, the amplitude of the distortion products is rarely more than 1 nm, at least in the 5−9 kHz frequency region addressed here. When the sensitivity of a preparation is reduced, the maximum distortion at a stimulus frequency of
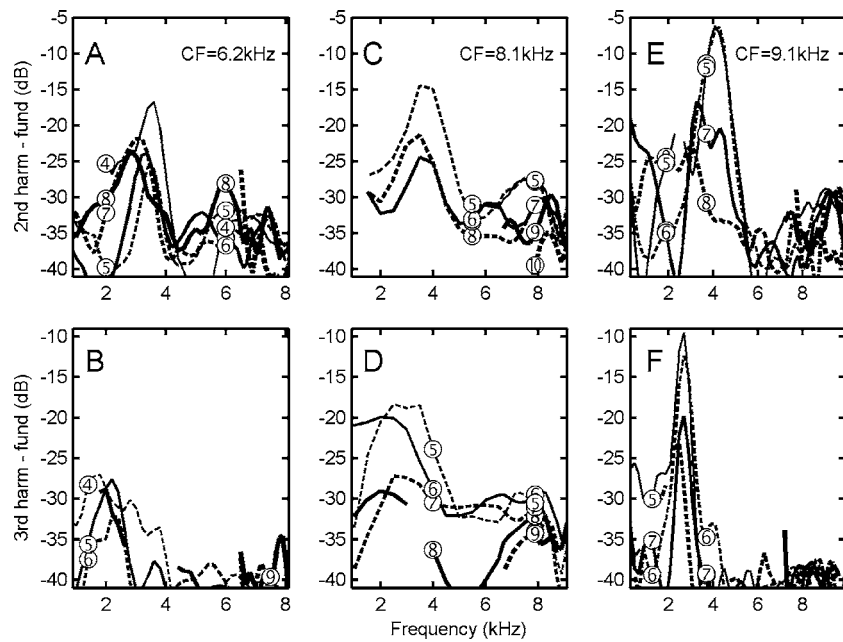
FIG. 9. Magnitudes of the second and third harmonics relative to the fundamental as a function of the frequency of the fundamental. Alternate SPLs are indicated with solid and dashed lines and the number symbols code intensity for 10 X's the number in the symbol. Data for which the magnitude was too small and therefore in the noise have been suppressed. Increasing SPLs are indicated by increasing thickness of the lines. A,B. set of data for a location with CF=6.2 kHz. N92. C,D. A second set of second and third harmonic data for location with the CF=8.1 kHz. E,F. A third set of data for a location with CF =9.1 kHz. N157.

CF/2 may be significantly reduced or even absent while the distortion amplitudes at CF can be larger than in high sensitivity cochleae.

The distortion amplitude does not vary much with increasing stimulus level and usually decreases relative to the fundamental at high stimulus levels. No data are shown for frequencies above the cutoff of the CF region as the amplitudes are too small to be reliable and likely do not contribute to any perception. When the SPL required to produce the second harmonic amplitude is determined from the CF-IO function, it is always less than 20 dB SPL. There is considerable variability in the amplitudes of the second and third harmonic as seen in comparing columns 1 and 2. While as a rule the second harmonic is the largest distortion component, in Fig. 9(d), the third harmonic is somewhat larger than that in Fig. 9(b) over the entire frequency range.

In the exceptionally sensitive preparation that is presented in Figs. 2 and 4, the maximum distortion products were exceptionally large, second at −6 dB and the third −9 dB re fundamental [Figs. 9(e) and 9(f)]. Another oddity is that the maxima are at frequencies lower than CF/n. Part of the shift could be due to the shift of the maximum with increasing intensity to lower frequencies. The relative amount of distortion decreases with intensities >60 dB SPL. The magnitude of distortion when the stimulus frequency is in the compressive region, around CF, is always less and usually <−35 dB re the fundamental.

G. Wave velocity

There are two aspects of the velocity of propagation of the TW along the length of the cochlear partition. The velocity of propagation of sound in a fluid is 1500 m/s which implies the acoustic signal in the cochlea essentially reaches the apex almost instantaneously. An example of the speed for which excitation begins in the cochlear partition is the latency of movement in response to a click. In response to a 25 μS click, the basilar membrane at this location, ~3.5 mm from the base of the cochlea, has a latency of 40 μS relative to that of the umbo [Fig. 10(a)]. There is also the customary modulation (Recio *et al.* 1998) of the BM impulse response which has been attributed to a second mode of vibration or reflections in the cochlea (Lin and Guinan, 2004; Shera and Guinan, 2003).

The response to a 10 μS click shown in Fig. 10(b) was obtained from a very sensitive preparation with a CF of 9.1 kHz (cf. Fig. 2). The remarkable aspect is the long duration that the BM oscillates up to, 25 ms and possibly longer. The modulation of the envelope at approximately 1 kHz persists for >20 ms.

The TW velocity is estimated by using the phase transfer functions obtained at two points along the cochlea and the relation, velocity=distance/time. There is a limitation to the range of frequencies that the velocity can be computed due to the fact that at low frequencies for a given location the phase difference between the two locations is nearly zero resulting in noisy differences in phase that often oscillate between large positive and negative values. Two examples of velocity measurements are shown in Fig. 11. They indicate that the traveling wave rapidly slows down as it approaches its characteristic location where viscous and transducer forces absorb the acoustic energy. The TW slows to around 2 m/s in each instance and is best shown in the insets. In a single instance the TW slowed to 1 m/s for a frequency above CF. The maximum velocity at low frequencies cannot be derived due to the noisy phase curves. However, it is clear that the velocity begins a rapid decrease as soon as the com-
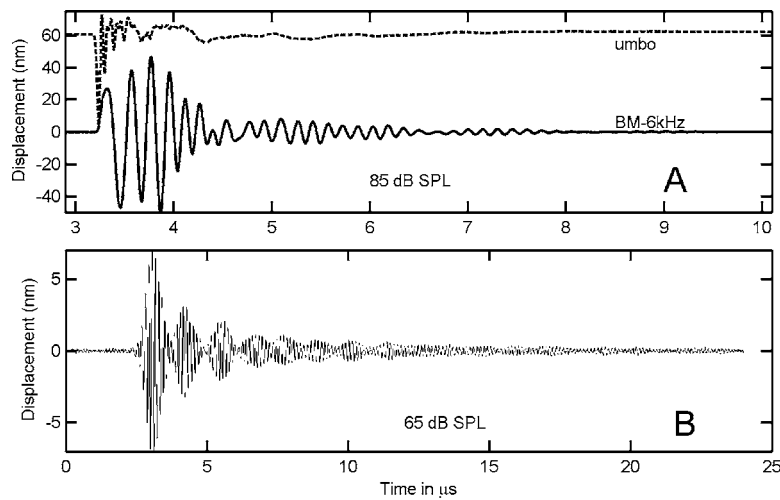
FIG. 10. A. Comparison of the umbo and BM (6 kHz) responses to a 25 μs click at 95 dB pSPL. The click occurred at 3 ms to avoid an equipment artifact that appears in very sensitive preparations at 0 s delay. The initial displacement of the umbo to the condensation click is away from the interferometer while the BM response is towards it. BM latency relative to that of the umbo is 40 μs. N136. B. The BM response to a 10 μs click of a very sensitive cochlea at 65 dB pSPL. Cochlea and cochlear location the same as that in Fig. 2. CF=9.1 kHz. N157.

pressive region for the CF is reached. The fact that the velocities calculated from these very small phase differences at low frequencies oscillate between + and − values suggests that the velocity is very fast and BM movement is likely in phase with the compressive wave and stapes velocity. That is, there is little delay in the BM following frequencies that are low relative to the CF of the region.

H. Osseus spiral limbus motion

There are at least three reasons for interest in whether the OSL vibrates; (1) the plateau region seen in BM transfer
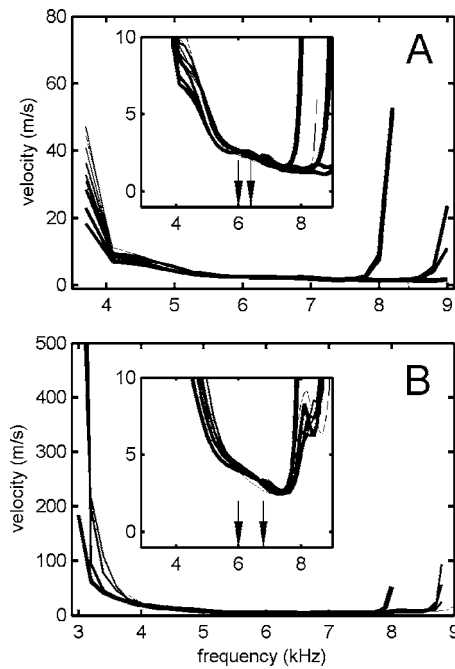
FIG. 11. Two examples of traveling wave velocity as it approaches the CF region where there is compression (nonlinearity). Velocity at increasing intensity is indicated by increasing line thickness. Insets have an expanded vertical scale and the arrows indicate the 6 kHz CF. A. Transfer functions were obtained at two locations separated by 190 μm with CFs=6.0 and 6.4 kHz. N136. B. Two locations separated by 300 μm with CFs=6.0 and 6.8 kHz. N92 (different CF location than used for the data in Fig. 1).

functions has not been seen in AN fiber tuning curves (discussed by Robles and Ruggero, 2001) though it has been observed with pressure and cochlear microphonic measurements (Olson, 1998; Cheatham and Dallos, 1997; DSchmiedt and Zwislocki, 1977); 2) the tip/tail ratio for the BM transfer functions appears to be smaller than typically found in AN tuning curves that were obtained in an unpublished study of chinchilla AN (from this lab) though they agree with those (30–40 dB) found in cat (Liberman, 1978); and 3) there is a suggestion in the present data that there is additional signal filtering as a result of motion of the OSL. The problem with comparing OSL motion to that of the BM at a given location is that the actual location of the measurement is likely of great importance. It requires "luck" to get two beads appropriately located for the comparison to be made properly, i.e., to have a reflective bead very near the edge of the OSL and also a second bead on the BM near the near the center of the BM. The further removed from the site at which the foot of the inner pillar rests the smaller the OSL vibration is.

Examples of the BM/OSL TFs in Fig. 12 indicate that, although the OSL does not vibrate much relative to that of the BM over the compressive frequency region, OSL vibration transfer functions reflect local BM vibration including nonlinearity near CF. However, the OSL vibrates with significant amplitudes relative to the basilar membrane at low frequencies, i.e., up to about one-half octave below CF, there can be less than a 5–15 dB difference. Below CF there is between a 3 and 7 dB slope to the BM/OSL transfer ratio that implies at some low frequency the ratio will be zero resulting in little or no excitation of the hair cells. Another factor is the relative phase between the OSL and the BM. While at low frequencies there is only a small phase difference between them, there is a continuously varying relation over the compressive frequency region where, due to the large amplitude ratio, the phase difference may not matter. Energy transfer in this frequency region must be from the BM to the OSL because the IO function is compressive. For frequencies outside this region, the energy transfer is likely direct from the pressures in the incompressible fluid of the
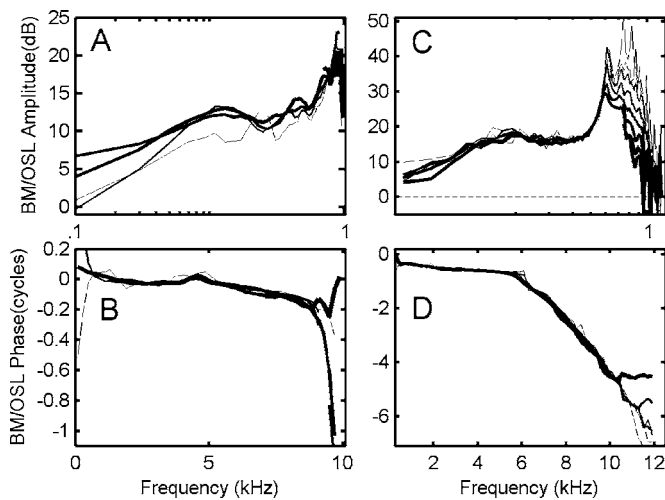
FIG. 12. A, C. Two examples of the vibration amplitude of the BM relative to that of the OSL. Thicker lines indicate higher intensities. N148. B, D. Phase transfer functions corresponding to the amplitude transfer functions in A and C respectively. N99.

cochlea (cf. Olson, 1998). In the plateau region the OSL/BM amplitude ratio is as little as −10 dB and there is a relatively constant phase difference which could be important for inner hair cell (IHC) excitation.

However, to obtain a more complete picture likely requires the measurement of vibration of the location of the attachment of the tectorial membrane to the OSL. At this site the OSL thins down substantially and could be vibrating enough to greatly reduce any excitation of the inner hair cell cilia.

IV. DISCUSSION

A. Goals of this study

Using the open bulla approach to the basal region of the cochlea, an attempt was made to obtain BM mechanical responses at the lowest CFs possible using a displacement-sensitive interferometer. While 5 kHz CFs were obtained on a few occasions, none of those preparations were very sensitive and as such the data are not included here. The best cochlear sensitivities were obtained for CFs between 6 and 9 kHz.

Several findings stand out from this study:

1) Extension of cochlear measurements to the 6 kHz region of the cochlea demonstrate that frequency tuning in this region is at least as good as those at more basal locations (higher CFs).
2) Detailed IO functions at small frequency steps yielded detailed frequency transfer functions that show what appears to be the interaction of a TW and a fast compression wave. With every 2π radian increase in the phase of the TW, the addition of a fast wave of amplitude equal to that of the TW at opposite phase results in a "null" which is seen as a notch in the IO functions and a ripple in the higher level frequency transfer curves.
3) Cochlear gain is better than 60 dB in sensitive cochleae and has been shown to approach 80 dB in one cochlea, a value larger than previously reported.
4) Rate of growth of BM vibration is at least as low as 0.1 dB/dB at CF, lower than previously reported.
5) A compressive cochlear nonlinearity is present at intensities >105 dB SPL and likely continues to higher intensities at CF.
6) In sensitive cochleae there is an expansive nonlinearity for frequencies in both the tail and the plateau frequency response region.
7) The amplitude of vibration of the OSL, while less than that of the BM, may be sufficient to play a significant role in reducing or possibly eliminating AN responses at low frequencies and could also affect IHC excitation in the plateau region.
8) There is good agreement between BM and AN threshold tuning curves, in agreement with prior studies (Ruggero *et al.* 2000).

Except for the 6 kHz studies in squirrel monkey, where the Mössbauer measurement technique was used (Rhode, 1971), and a single study by Rhode and Recio (2000), little data have been reported for mid-frequency CFs (<8 kHz). Overall the response properties of the basilar membrane reported here are compatible with and an extension of those catalogued in the review of Robles and Ruggero (2001). BM vibration in this region has sensitivity that can be greater than 90 dB re 1 nm/Pa, and a single instance of 96 dB sensitivity corresponding to a displacement sensitivity of 63 μm/Pa was observed, greater than any previously reported (Table 1 of Robles and Ruggero, 2001). This level of sensitivity is remarkable, in that if the cochlea did not possess a strong compressive nonlinearity, the BM vibrations at 100 dB SPL would be 60–120 μm and would certainly result in the destruction of the tip links of the stereocilia thereby eliminating forward transduction in the OHCs. These high sensitivities indicate the compressive attribute of the cochlea is absolutely necessary to preserve its integrity. The observed BM operating range in sensitive cochleae is 1–200 nm for intensities less than 100 dB SPL and is in accord with reports of the operating range of the hair cells' mechanoelectric transducer, 1–250 nm (e.g., Fettiplace and Ricci, 2006).

In several experiments, the compressive nonlinearity is evident at 0 dB SPL though the more typical result is that it starts near 20 dB SPL. Coupled with this is the fact that the IO function at CF usually does not achieve linearity at high levels, 100–105 dB SPL, or even at 115 dB SPL when tested in a single experiment, indicating that gains calculated between 0 and 80 dB SPL are likely to be low estimates (note that in over 50% of the present studies that the slope of CF-IO functions increases between 90 and 100 dB SPL and trends toward linearity possibly because the preparation is not of high sensitivity). In the present experiments, gains of 60 dB were found for intensities less than 100 dB SPL (7 of 40), five cochleae had gains >65 dB, and, as shown in Fig. 7(b), a gain of nearly 80 dB was obtained in a single instance. This is a truly remarkable achievement in biological design that remains to be explained.

Estimating the gain of the cochlear amplifier is somewhat problematic. While there is an indication that 60 dB

gain is a reasonable amount under the best of conditions (Robles and Ruggero, 2001), the data presented here indicate that it can be greater. The procedure used to estimate the gain can be either the horizontal shift from the CF-IO curve to the asymptotic linear curve or the difference in sensitivity at CF between threshold and some maximum stimulus level. In the latter case, depending on the level selected one could obtain different gains. In addition, with nonlinearity beginning below 0 dB SPL, what should be the reference level? There are limitations in the reported data because often when collecting data one may not be able to cover the full intensity range from 0 up to 100 or 105 dB SPL due to sound-source or measurement device limitations. Also, when using the IO-horizontal shift technique it is not always apparent at what intensity that the IO curve becomes linear, if at all, at high levels (e.g., Fig. 7). Based on the present measurements, compression at CF in the 6 kHz region can continue to arbitrarily high levels ($>$115 dB SPL). There was no indication of the IO function becoming linear at any intensity ($<$100 dB SPL) in either squirrel monkey (Rhode, 1971) or chinchilla (Robles et al., 1986) when the Mössbauer measurement technique was used.

There are two possible mechanisms to explain cochlear gain that are known to have active properties: (1) electromotility due to the prestin molecule in the wall of the OHCs (Zheng et al., 2000), and (2) the tip-link adaptation mechanism in the cilia (Martin and Hudspeth, 1999; Fettiplace et al., 2001; Ricci et al., 2002). Both mechanisms have been shown to have fast kinetics that could operate over the entire frequency range. It has been argued that, because the basal lateral aspect of the OHCs is effectively a low-pass filter with a 1 kHz corner frequency, the intracellular ac voltage will be attenuated such that there would never be sufficient "voltage drive" for the electromotility to work at high frequencies. There are proposals that attempt to avoid this limitation that involve feedback from the OHCs and/or their piezoelectric properties (Mountain and Hubbard, 1989; Liao et al., 2005). Or perhaps cochlear microphonic potentials which are present throughout the cochlea could provide the necessary voltage drive for electromotility (Dallos and Evans, 1995). Cochlear amplification might be achieved by summating the effects of a group of OHCs so that individual cells do not have to be fast (Lu et al., 2006). It may be possible that the OHC electromotility is involved in a slow process that adjusts the operating point of the OHCs, while the tip-link motors of the stereocilia provide the fast adaptation necessary for a high frequency gain mechanism (e.g., Kim, 1986).

An aspect of cochlear mechanics that is poorly understood is what happens in the low frequency linear region of the response where a 100 dB SPL stimulus would result in 1 μm displacements. A 1 μm displacement would likely be sufficient to cause hair cell damage and 100 dB SPL is by no means the most intense stimulus that individuals are exposed to. However, there are several other peripheral gain-control mechanisms that affect the low frequency magnitudes: the stapedius and the tensor tympani both attenuate the middle-ear responses (e.g., Møller, 2000). Olivo-cochlear feedback may further attenuate the motion of the basilar membrane (Dolan and Nuttall, 1994). Yet another mechanism that could reduce the displacement of the stereocilia is motion of the OSL resulting from the fast-compression wave. The relative motion between the reticular lamina in which the apical ends of hair cells are embedded and the tectorial membrane in which the stereocilia of the OHCs are embedded would be attenuated by any comparable in-phase motion of the OSL. The OSL displacement that one would like to determine is where the foot of the inner pillar cell rests on the OSL as this is a key location for determining the rotation of the reticular lamina. The further away from this location that measurements are made, the smaller the vibration amplitudes. Because of the role of the tectorial membrane in determining displacement of the stereocilia of the OHCs, its vibration, and the vibration of its attachment to the OSL play a key role in the final excitation of the IHCs and as such need to be studied. A recent study using electrical stimulation of the OHCs along with measuring the motion of the apical surface of the IHCs and the opposing tectorial membrane demonstrated that there is direct fluid coupling between the OHCs and the IHCs sufficient to move the IHC stereocilia, at least for frequencies $<$3 kHz (Nowotny and Gummer, 2006).

When nonlinear IO functions become linear as intensity is increased it likely results from the linear fast wave becoming the dominant drive on the basilar membrane. The level at which that occurs is dependent on the relative amplitudes of the fast wave and the slower traveling wave. The TW has larger amplitude at low levels than the fast wave. Once the two are of equal amplitudes the interaction between the two is expressed as a notch in the IO function where the phase changes rapidly (cf. Fig. 6). In fact, a phase change of 180° for a 1 dB level change has been recorded (cf. Fig. 6). The intensity at which the notch occurs decreases as the frequency increases above CF [Fig. 5(a)], a result of the decrease in TW amplitude. The final asymptotic phase of the BM vibration varies in multiples of 2π as stimulus level is increased [Fig. 1(d)]. This correlates with TW phase changing by $\sim2\pi$ with increasing frequency and the phase of the fast wave remaining relatively constant allowing cancellation and hence notch/ripple formation. Notably, constant phase implies that the vibration has a very high velocity implying high fluid velocity that could arise from either a compression or an evanescent wave. Constant phase indicates that the fast wave dominated the response at these frequencies and intensities.

B. Harmonic distortion

Second harmonic distortion is the largest at a site tuned to CF when the stimulus frequency is CF/2, and third harmonic distortion is largest when the stimulus frequency is CF/3. The relative magnitudes of the distortions are usually less than −20 dB re the fundamental. There were occasionally instances when the distortion was −15 dB and in the most sensitive preparation it reached −6 dB (50%) relative to the fundamental. The growth of distortion is nonlinear and nonmonotonic, attaining a maximum at some intermediate intensity level. Overall, the results are largely compatible with those obtained by Cooper (1998) in guinea pig in the 17–18 kHz region.

One reason to be concerned with this result is due to the studies of Martin *et al.* (1999). They found that as an interference tone, $f3$, is swept through a frequency range that includes the second harmonics of the primary frequencies, $2f1$ and $2f2$, there are combinations of $f3$ and the harmonics that result in distortion products (DPs) that can sum with those generated by $f1$ and $f2$. Present data suggest that there could be distortion components of sufficient magnitude to result in a distortion product otoacoustic emission (DPOAE) being generated at the high frequency site that would explain their results.

C. Osseous spiral OSL motion

The earliest report of OSL vibration indicated that the BM/OSL amplitude ratio was >20 dB at low frequencies and increased in the CF region in squirrel monkey (Rhode, 1971, Fig. 10). There were no data shown in the plateau region, and it is likely that these initial data were not obtained in the most sensitive cochleae.

It is somewhat surprising that the OSL moves as much as it does at low frequencies as it is a massive structure relative to that of the BM in the basal region of the cochlea. Movement of the OSL would act as a second filter in the excitation of the inner hair cells. The OSL transfer functions were somewhat erratic and defy an exact description. However, the OSL at low frequencies in the high-CF region of chinchilla cochleae can move with nearly the same amplitude and phase as the basilar membrane. In several instances the low frequency slope of the OSL/BM function is on the order of a few dB/oct (average=3.3 dB/oct) implying nearly the same amplitude for both structures and hence reduced excitation of the hair cells. The principal effect is for frequencies below a few hundred Hz where there would be significant attenuation of IHC excitation. One of the unexplained differences between AN and BM tuning how a low frequency tone can suppress the rate output of an AN fiber when it is known that the peak of the basilar membrane displacement of the combination of the two tones is always greater than the amplitude produced by the CF tone alone (Cooper, 1996). Phase differences between the BM and OSL vibration could reduce tectorial membrane motion relative to that of the reticular lamina thereby reducing IHC excitation.

Attempts to measure OSL motion in the plateau region of vibration for a given CF were hampered by the need to use 100 dB SPL stimuli to obtain reliable measurements, but that was not always possible for frequencies >10 kHz. IO functions verified that the region was responding in a linear manner. At times it appeared that OSL motion was sufficient to account for the lack of IHC excitation in the plateau region though usually there was at least a 10 dB difference in the amplitudes of the OSL and the basilar membrane. By contrast, the BM/OSL ratio at 1 kHz was ~10 dB on average.

Békésy (1949) noted that in his human temporal bone measurements the central axis of the cochlea (read as spiral OSL) may move near the helicotrema. He also stated that 7–8 mm of the BM surrounds the helicotrema, a feature that is similar in chinchilla. Recently, Stenfelt *et al.* (2003) reported that their measurements in human temporal bone that the OSL vibrated with the same amplitude as the BM.

It is likely that OSL motion in the plateau region affects inner hair cell excitation. However, it is possible that the vibration of the tectorial membrane with a different resonant frequency than the basilar membrane could reduce excitation in this frequency region (Zwislocki and Kletsky, 1979; Allen, 1980).

D. Some conclusions

The present results indicate that in sensitive cochleae that nonlinearity is present throughout the frequency region to which a particular location responds. This is in contrast to the traditional view that IO functions in both the tail and plateau region are strictly linear. The occurrence of the ripple pattern in the frequency transfer functions may be explained by combining the response to the traveling wave with a fast, compression wave. The ripple pattern in iso-intensity frequency functions indicates that the two waves interact over the entire frequency range (esp., Fig. 4). Though it is speculative that the resulting "ripple pattern" in the tail region of the response results from this combination.

Prior measurements and modeling efforts suggest that there are possibly many modes of vibration of the cochlear partition (Hubbard and Mountain, 1996, Nilsen and Russell, 2000). Given the complex cochlear anatomy there is that possibility. However, there are reports that the vibration of the BM along the radial direction is "simple," i.e., without modes (Cooper, 1999; Rhode and Recio, 2000). The present data addressing OSL vibration suggest that with some further refinement of the measurements, the contribution of OSL motion and the fast compression wave may explain all aspects of AN fiber behavior.

ACKNOWLEDGMENTS

Allen, J. B. (**1980**). "Cochlear micromechanics-A physical model of transduction," J. Acoust. Soc. Am. **68**, 1660–1670.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects," J. Acoust. Soc. Am. **49**, 1131–1139.

Von Békésy, G. (**1949**). "The vibration of the cochlear partition in anatomical preparations and in the models of the inner ear," J. Acoust. Soc. Am. **21**, 233–245.

Carr, C. D., and Bohne, B. A. (**1979**). "Location of structurally similar areas in chinchilla cochleas of different length," J. Acoust. Soc. Am. **66**, 411–414.

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

Cooper, N. P. (**1998**). "Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea," J. Physiol. (London) **509**, 277–288.

Cooper, N. P. (**1999**). "Radial variation in the vibration of the cochlear partition," in *Recent Developments in Cochlear Mechanics*, World Scientific, Singapore, pp. 109-115.

Cooper, N. P., and Rhode, W. S. (**1992**). "Basilar membrane mechanics in the hook region of cat and guinea pig cochlea: Sharp tuning and nonlinearity in the absence of baseline shifts," Hear. Res. **63**, 163–190.

Cooper, N. P. (**2000**). "Radial variations of the cochlear partition," In *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 109–115.

Cheatham, M. A. and Dallos, P. (**1997**). "Intermodulation components in inner hair cell and organ of Corti responses," J. Acoust. Soc. Am. **102**, 1038–1048.

Dallos, P., and Evans, B. N. (**1995**). "High frequency motility of outer hair cells and the cochlear amplifier," Science **167**, 2006–2009.

Dolan, D. F., and Nuttall, A. L. (**1994**). "Basilar membrane motion evoked by sound is altered by electrical stimulation of the crossed olivocochlear bundle," Assoc. Res. Otolaryngol. Abstr. **17**, p. 89.

Fettiplace, R., and Ricci, A. J. (**2006**). "Mechanoelectrical transduction in auditory hair cells," in *Vertebrate Hair Cells*, edited by R. A. Eatock, R. R. Fay, and A. Popper, **SHAR 27** (Springer, New York).

Fettiplace, R., Ricci, A. J., and Hackney, C. M. (**2001**). "Clues to the cochlear amplifier from the turtle ear," Trends Neurosci. **24**, 160–175.

Goldstein, J. L. (**1995**). "Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane and MBPNL model," Hear. Res. **89**, 52–68.

Hubbard, A. E., and Mountain, D. C. (**1996**). "Analysis and synthesis of cochlear mechanical function," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay, **SHAR 6** (Springer, New York), pp. 62–120.

Khanna, S. M., and Hao, L. F. (**1999**). "Nonlinearity in the apical turn of living guinea pig cochlea," Hear. Res. **135**, 89–104.

Kiang, N. Y. S., Liberman, M. C., Sewell, W. F., and Guinan, J. J. (**1986**). "Single unit clues to cochlear mechanisms," Hear. Res. **22**, 171–182.

Kim, D. O. (**1986**). "Active and nonlinear cochlear biomechanics and the role of outer-hair-cell subsystem in the mammalian auditory system," Hear. Res. **22**, 105–114.

Liao, Z., Popel, A. S., Brownell, W. E., and Spector, A. A. (**2005**). "Effect of voltage-dependent membrane properties on active force generation in cochlear outer hair cell," J. Acoust. Soc. Am. **118**, 3737–3746.

Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–445.

Lin, T., and Guinan, J. J., Jr. (**2004**). "Time-frequency analysis of auditory-nerve fiber and basilar membrane click responses reveal glide irregularities and non-characteristic-frequency skirts," J. Acoust. Soc. Am. **116**, 405–416.

Lu, T. K., Zhak, S., Dallos, P., and Sarpeshkar, R. (**2006**). "Fast cochlear amplification with sloe outer hair cells," Hear. Res. **214**, 45–67.

Martin, G. K., Barden, B., Stagner, J. 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.

Martin, P., and Hudspeth, A. J. (**1999**). "Active hair-bundle movements can amplify a hair cell's response to oscillatory mechanical stimuli," Proc. Natl. Acad. Sci. U.S.A. **96**, 1218–1231.

Møller, A. R. (**2000**). *Hearing. Its Physiology and Pathophysiology.*, Academic Press, New York, pp. 345–370.

Morioka, I., Reuter, G., Reiss, P., Gummmer, A. W., Hemmert, W., and Zenner, H. P. (**1995**). "Sound-induced displacement responses in the plane of the organ of corti in the isolated guinea-pig cochlea.," Hear. Res. **83**, 142–150.

Mountain, D. C., and Hubbard, A. E. (**1989**). "Rapid force production in the cochlea," Hear. Res. **42**, 195–202.

Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A. (**1998**). "Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae," Science **282**, 1882–1884.

Nilsen, K. E., and Russell, I. J. (**2000**). "The spatial and temporal represen-tation of a tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **22**, 11751–11758.

Nowotny, M., and Gummer, A. W. (**2006**). "Nanomechanics of the subtectorial space caused by electromechanics of cochlear outer hair cells," Proc. Natl. Acad. Sci. U.S.A. **103**, 2120–2125.

Nuttall, A. L., and Dolan, D. F. (**1996**). "Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig," J. Acoust. Soc. Am. **99**, 1556–1565.

Nuttall, A. L., Ren, T., Gillespie, P., Grosh, K., and de Boer, E. (**2006**). *Discussion Session. Auditory Mechanisms. Processes and Models* (World Scientific, Singapore), pp. 521–543.

Olson, E. S. (**1998**). "Observing middle and inner ear mechanics with novel intracochlear pressure sensors," J. Acoust. Soc. Am. **103**, 3445–3463.

Peterson, L. C., and Bogert, B. P. (**1950**). "A dynamical theory of the cochlea," J. Acoust. Soc. Am. **22**, 369–381.

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.

Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.

Rhode, W. S., and Cooper, N. P. (**1996**). "Nonlinear mechanics in the apex of the chinchilla cochlea in vivo," Aud. Neurosci. **3**, 101–121.

Rhode, W. S., and Recio, A. (**2000**). "Study of the mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.

Rhode, W. S., and Smith, P. H. (**1985**). 'Characteristics of tone-pip response patterns in relationship to spontaneous rate in auditory nerve fibers," Hear. Res. **18**, 159–168.

Ricci, A. J., Crawford, A. C., and Fettiplace, R. (**2002**). "Mechanisms of active hair bundle motion in auditory hair cells," J. Neurosci. **22**, 44–52.

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. (**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., Narayan, S. S., Temchin, A. N., and Recio, A. (**2000**). "Mechanical bases of frequency tuning and neural excitation at the base of the cochlea: Comparison of basilar-membrane vibrations and auditory-nerve-fiber responses in chinchilla," Proc. Natl. Acad. Sci. U.S.A. **97**, 11744–11750.

Schmiedt, R. A., and Zwislocki, J. J. (**1977**). "Comparison of sound-transmission and cochlear-microphonic characteristics in Mongolian gerbil and guinea pig," J. Acoust. Soc. Am. **61**, 133–149.

Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). "Measurement of basilar membrane motion in the guinea pig using the Mossbauer technique," J. Acoust. Soc. Am. **72**, 131–141.

Sewell, W. F. (**1984**). "The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats," Hear. Res. **14**, 305–314.

Shera, C., and Guinan, J. J., Jr. (**2003**). "Stimulus-frequency-emission delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.

Stenfelt, S., Puria, S., Hato, N., and Goode, R. L. (**2003**). "Basilar membrane and osseous spiral lamina motion in human cadavers with air and bone conduction," Hear. Res. **181**, 131–143.

Zheng, J., Shen, W., He, D. Z., Long, K. B., and Dallos, P. (**2000**). "Prestin is the motor protein of cochlear outer hair cells," Nature (London) **405**, 149–155.

Zwislocki, J. J., and Kletsky, E. J. (**1979**). "Tectorial membrane: A possible effect on frequency analysis in the cochlea," Science **204**, 639–641.

Mutual suppression in the 6 kHz region of sensitive chinchilla cochleae

William S. Rhode[a)]

Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706

Basilar membrane (BM) vibration was measured using a displacement measuring interferometer for single-tone and two-tone suppression (2TS) paradigms in the 6–9 kHz region of sensitive chinchilla cochleae that had gains near or better than 60 dB. Based on prior studies of basilar membrane vibration, three significant differences remain between BM and auditory nerve (AN) 2TS responses: (1) suppression thresholds in the tail of tuning curves were much higher in BM than the auditory nerve (AN); (2) rates of suppression were significantly higher in AN than BM; and (3) the amplitude of vibration with low-frequency suppressors was always greater than the single-tone displacement rendering it impossible to explain 2TS rate suppression in the AN. The first two differences are eliminated by the results of the present study while the third remains. Suppression amplitudes greater than 40 dB and rates of suppression larger than 2.5 dB/dB were found for low-frequency suppressors. A correlation between both the gain and nonlinearity of the cochlea and 2TS properties indicates that when sensitive cochleae are studied. The third difference between BM and AN behavior could be strictly a function of the high-pass filter characteristic of the inner hair cells. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2718398]

I. INTRODUCTION

Two-tone suppression (2TS) is a result of the nonlinear nature of cochlear transduction: when the nonlinearity vanishes so does 2TS (Ruggero *et al.*, 1992). Two-tone suppression has been studied in cochlear mechanics (Ruggero *et al.*, 1992; Rhode and Cooper, 1993; Cooper, 1996), the auditory nerve (e.g., Schmiedt, 1982; Fahey and Allen, 1985; Delgutte, 1990), hair cells (Sellick and Russell, 1979; Cheatham and Dallos, 1992; Nuttall and Dolan, 1993) and models (e.g. Zwicker, 1979; Geisler and Nuttall, 1997). The degree to which a cochlea exhibits this nonlinear transduction appears to be dependent on the state of outer hair cells (OHCs). OHCs have been demonstrated to be the key element in attaining the exquisite sensitivity of the auditory system, compression of response amplitudes, and generation of distortion products (Robles and Ruggero, 2001). However, it is not yet agreed upon whether it is the operation of the OHC cilia (Hudspeth and Jacobs, 1979; Chan and Hudspeth, 2005) and/or the electromotile property of the prestin-based molecular motors in the walls of OHCs (Zheng *et al.*, 2000) that is responsible for this nonlinearity.

Due to constraints of anatomy, i.e., the cochlea being a tightly coiled helical structure largely embedded in bone, the study of cochlear mechanical suppression has been largely relegated to the basal cochlear region, mostly in rodents. While "log scaling" of basal cochlear measurements is a convenient manipulation to picture the spatial distribution of cochlear phenomena, it is important to perform studies across the entire cochlea to verify this scaling assumption, and particularly in the cochlear region where speech is principally encoded. This is a challenging task that has yet to be fully realized. Log scaling breaks down for frequencies <1 kHz (Greenwood, 1990) so the present results obtained in the 6–9 kHz region of the chinchilla cochlea are in the log-scaling region. Chinchillas have a similar perceptual range of frequencies to humans (Fay, 1988), though its cochlear length is approximately one-half that of humans (Carr and Bohne, 1979). Because hair cell diameters are relatively constant across species, this latter characteristic implies that chinchillas have approximately half the hair cells per Hz that humans have which may result in some species differences in cochlear response properties.

Prior studies have illustrated both the amplitude and phase responses of 2TS. The magnitude of suppression is correlated with the sensitivity of the preparation. Though there is much that the vibration of the basilar membrane (BM) has in common with other cochlear response measures, there remain differences that require further explanation. There are three principal characteristics of neural 2TS that apparently differ from BM-2TS: (1) differences in suppressor threshold levels, (2) differences in rate of suppression (ROS), and (3) the neural response is typically lower than the response to the probe alone, unlike the BM response (Cooper, 1996; Geisler and Nuttall, 1997; Sen and Allen, 2006). The present study addresses cochlear 2TS in general and eliminates the first two apparent discrepancies.

II. METHODS

Methods are essentially those detailed in Cooper and Rhode (1992). Sixteen chinchilla cochleae were studied between 3.5 and 4.25 mm from the basal end of the basilar membrane where the characteristic frequency (CF) is between 5.3 and 9.1 kHz. CF is defined as the frequency that

---
[a)]Electronic mail: rhode@physiology.wisc.edu

yields the largest response at threshold. All procedures were approved by the Animal Care and Use Committee of the University of Wisconsin.

Each animal was anesthetized with pentobarbital using a dose rate of 75 mg/kg. Additional doses were administered to maintain the animal in a deeply areflexive state. All anesthetics were administered intraperitoneally. A tracheotomy was performed to ensure an open airway and to place the animal on a respirator if necessary, though it was never used in this set of experiments. After the ear was surgically removed, four screws were implanted in the skull and cemented in with dental cement in order to form a rigid base. A bolt was then cemented to the base to provide a stable fixation of the skull to a head holder with six degrees of freedom for the purpose of positioning the cochlea under the microscope.

No compound action potentials were recorded and no middle-ear muscles were cut in order to minimize any possible traumatic effects of preparing the cochlea. The sensitivity of the preparation at CF was measured by repeated BM input-output (IO) functions at CF and near the center of the BM. In a few instances sensitivity was maintained over a 12 h period. Measurements were terminated if there was more than a 6–8 dB decrease in the sensitivity at CF. In every case where the sensitivity was high, measurements were made near the center of the BM.

The overlying cochlear bone in a region with characteristic frequencies (CFs) between 6 and 9 kHz was shaved down using a microdrill with a 0.45 mm burr until the remaining tissue and/or bone debris could be removed with a pick fabricated out of a microelectrode and/or the use of a stream of artificial perilymph to flush the debris. Gold-coated polystyrene beads 25 μm in diameter served as retroreflectors. They were placed in the perilymph and allowed to sink to the basilar membrane. They have a specific gravity of 1.05, which is near that of water (1.0) and therefore any loading of the basilar membrane by the bead should be minimal. To support that belief, Cooper (1999) showed that there are no detectable differences between the bead and no-bead measurements of BM vibration. In nearly all the preparations reported here, the bead was located near the center of the BM. A glass cover slip was placed over the cochlear opening with no hydromechanical seal. The cover glass served to avoid the problem of an unstable air-fluid interface.

A plastic tube was inserted into the external ear canal and cemented in. The tube coupled both the source and allowed insertion of calibrated probe tube to which a 1/2 in. Bruel and Kjaer condenser microphone is adjoined. The tip of the probe tube could be visualized through an opening in the bony external ear as it was positioned parallel to the tympanic membrane within 1 ml of the tip of the malleus. The opening was sealed with a glass cover after a 45 μm bead was placed on the tympanic membrane at the tip of the malleus (or umbo). The bead was used as a retroreflector for the interferometer and allowed the measurement of the transfer function of the malleus. Two sound sources were used, an Etymotic ER10C phone which was also used to acoustically calibrate the delivery system, or a pair of Schure 5C phones

which were calibrated using the B and K probe tube microphone. Stimuli were presented using TDT System I and purpose-built software.

A. Single-tone experiments

Basilar membrane input-output (IO) functions were determined using 30 ms tone bursts with 1 ms raised cosine rise and fall times, presented every 100 ms. Stimulus amplitude typically covered a 100 dB SPL range in 5 dB sound pressure level (SPL) steps. Eight-to-ten responses were averaged for each stimulus condition. Analysis consisted of Fourier decomposition of the steady-state portion of the averaged response at the stimulus frequency after a Hamming window function was applied. Frequency transfer functions were constructed from the IO functions.

B. Two-tone experiments

In two-tone suppression (2TS) experiments, the two primary tones are designated fs and fp here, where fs and fp are the frequencies of the suppressor and probe tones, respectively. The fp is usually set to the CF of the cochlear region under study. Each of two primary tones was presented through separate phones coupled with plastic tubing to the earpiece. Both tones were 50 ms in duration and repeated every 175 ms with the "suppressor tone" delayed 10 ms in onset relative to the probe tone. Each tone had 1 ms rise/fall times and the sequence was repeated twice. The intensity range was typically 20–80 dB SPL with either 5 or 10 dB steps. Typically, the suppressor frequency range was approximately from one-half octave below to one-half octave above the CF in steps of 300 Hz. This range was focused on because it is frequently used in the study of the cubic difference tone in otoacoustic emission measurements. In a few experiments, a larger frequency range was covered that included several octaves lower than CF to an octave above CF.

C. Recording system

Mechanical responses were measured using a custom-built displacement-sensitive heterodyne laser interferometer (Cooper and Rhode, 1992). The laser was coupled to the preparation using a long working distance lens (Nikon SLWD 5X, NA 0.1). The laser was focused to an area of ~5 μm diameter on the reflective beads. The interferometer was not sensitive enough to measure basilar membrane vibration without the gold-coated beads. Instantaneous phase was measured using two single-cycle phasemeters that worked in quadrature. The phasemeter outputs were sampled at 250 kHz using a 16 bit analog-to-digital (A/D) converter and the phase was unwrapped using custom software. Response amplitudes were corrected for the frequency response of the recording system. The noise floor was <5 pm/$\sqrt{Hz}$ which allows measurements down to 0.01 nm for frequencies >6 kHz.
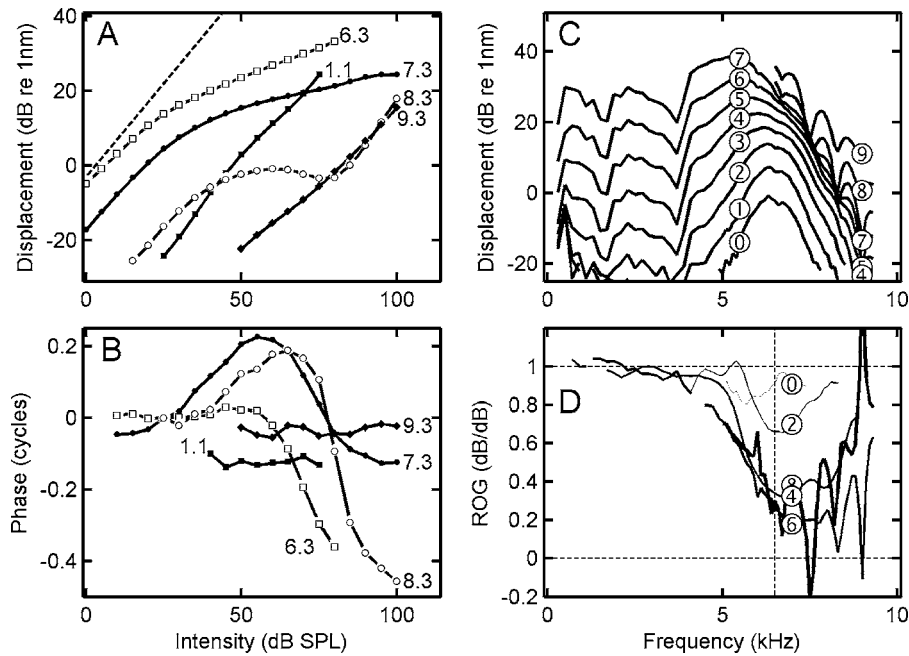
FIG. 1. Single tone responses for the 6 kHz region in a sensitive cochlear preparation. A. Input-output functions for the CF=6.3 kHz cochlear region. The dashed line indicates linearity. B. Phase IO functions corresponding to the functions in A. Where applicable phase-frequency functions were normalized by the 25 dB SPL phase function in order to emphasize the frequency-intensity dependence. The 9.3 and 1.1 kHz curves are positioned arbitrarily along the ordinate for clarity, the point being that there is no change in phase with increasing intensity. C. Frequency transfer curves at the intensities indicated by the numeric symbols. (e.g., ②=(dB SPl/10)=20 dB SPL). D. ROG: rate of growth in dB/dB. The slope of the IO functions as a function of frequency and intensity. The functions for SPLs <60 dB were three point smoothed using filtfilt (MATLAB). Only every fourth function, i.e., those that are separated by 20 dB SPL intensities, is shown for clarity. Data were collected as IO functions at 68 frequencies: in 200 Hz steps below 4.9 kHz and 100 Hz steps above; intensity: 0–100 dB SPL in 5 dB steps, stimuli were 30 ms in duration every 100 ms with five repetitions. Highest level attained at any frequency was limited by the maximum amplitude that could be recorded with the interferometer. Data corresponding to amplitudes <−26 dB (0.05 nm) were considered noise and eliminated from the graphs. Chinchilla N92.

III. RESULTS

A. Single-tone basilar membrane transfer functions

Five of the 68 IO functions for a single 6.5 kHz cochlear region illustrate the frequency dependent nature of the cochlear nonlinearity in Fig. 1(a). Frequencies <CF/2 (e.g., 1.1 kHz) have linear IO functions; frequencies closer to CF (e.g., 6.3 kHz) start to show compression at lower and lower intensities as CF is approached; at CF the rate of growth (ROG) of motion can be as low as 0.1 dB/dB; and beyond CF, rate of growth can approach 0 dB/dB. There are frequencies >CF where the IO functions have notches, i.e., nonmonotonicities such as seen in the 9.3 kHz function, and the slopes can be <0 and >1. These notches are thought to reflect the interaction of two modes of vibration in the cochlea (e.g., Lin and Guinan, 2004). The corresponding phase IO functions have a similar behavior to that previously reported (Anderson *et al.*, 1971); i.e., with increasing intensity the phase leads below CF, little change at CF, and they have lags above CF [Fig. 1(b)]. However, in chinchilla this type of phase behavior holds only for intensities below 60 dB SPL. Above 70–80 dB SPL, only phase lags are seen, independent of the frequency. Another exception occurs at notch frequencies where there are 180° abrupt phase shifts (e.g., Rhode and Recio, 2000; Rhode, 2007).

The 68 IO functions for this cochlea are plotted as iso-intensity functions in Fig. 1(c). For low intensities, e.g., 0–10 dB SPL, and taking 1 nm as the threshold BM dis-

placement corresponding to threshold activation of the hair cells, the effective frequency range is equivalent to a spatial expanse of ~300 μm of the cochlea. With increasing intensity, linear growth is noted below CF, compressive growth around and above CF, with sharp minima at various frequencies in the higher intensity functions for frequencies above CF. The minima occur with a frequency spacing of ~800 Hz which may relate to the interaction of two vibration modes that have slightly different resonant frequencies and therefore can sum in either a constructive or destructive manner.

Perhaps it is easier to understand the complex ROG of frequency-specific BM vibration with a plot of the slopes of the IO functions [Fig. 1(d)]. ROG=1 dB/db=a linear response for frequencies low relative to CF. In the region of nonlinear growth, ROG is a complex function of both frequency and intensity. At CF the ROG is already <1 by 20 dB SPL and decreases to 0.3 dB/dB at higher intensities. Beyond CF, ROG continues to decrease and can approach 0 dB/db over restricted regions. At certain frequencies, ROG functions reflect the presence of notches in IO functions: that is, regions where ROG is <0 and can also be >1 dB/db indicating an expansive nonlinearity. The presence of notches has only been observed in sensitive preparations and therefore it is presumed not to be an artifact. In very sensitive cochleae, ROG can be <0.1 dB\dB and the difference in sensitivity at CF between the active and passive cochlear states is ~60 dB (Rhode, 2007).

FIG. 3. Mutual suppression example. A. Suppression of a 6.55 kHz probe tone response at 35 dB SPL by the suppressor tone as a function of its frequency and intensity. Numeric symbols indicate Ls in dB SPL/10 (e.g., ③=30 dB SPL). Dashed lines without symbols correspond to intervening 5 dB intensities. Curves were normalized by the probe displacement when Ls was 20 dB SPL at which there was no noticeable suppression. B. The effect of the probe tone on a 35 dB SPL suppressor tone response as a function of suppressor frequency and probe intensity. The arrow indicates increasing probe intensity as does increasing line weight. Data corresponding to suppressor amplitudes <0.1 nm were considered noise and were eliminated. Numeric symbols indicate Lp in dB SPL/10 (e.g., ④=40 dB SPL). N92.

FIG. 2. Dual sets of displacement-IO functions from a two-tone study. The probe tone (fp: solid lines) was set to CF (6550 Hz) and the suppressor tone (fs: dashed lines) was varied between 4500 and 9000 Hz in steps of 300 Hz (16 frequencies). The tone intensities were varied between 20 and 80 dB SPL in steps of 5 dB (169 combinations). Intensity of the suppressor tone (Ls) is indicated on the abscissa and the intensity of the probe tone (Lp) is indicated by the thickness of the lines with the thickest line corresponding to 80 dB SPL and thinnest to 20 dB SPL. The interaction of four suppressor frequencies with the probe is illustrated for $fs < fp$, fs near fp, and $fs > fp$. Tone pips were 100 ms repeated after 175 ms intervals, the suppressor tone was delayed 20 ms relative to the probe tone: two repetitions. Chinchilla N92, measured at a different cochlear location than for data in Fig. 1.

B. Mutual suppression

Two tones applied to any nonlinear system will result in suppression of one tone by the other. Either of the two "primary" tones can serve as the suppressor or probe tone. Often in 2TS studies the probe tone is set to the CF of the cochlear region under study at a single intensity just above the threshold for the region. In the present studies the intensity of the probe tone was varied over the same intensity range as the suppressor tone to obtain a more thorough representation of the nonlinear interaction between the two tones.

With the probe tone frequency (fp) set at ~CF(6550 Hz), the response amplitudes for both tones at four fp/fs ratios (of 16 obtained) are shown in Fig. 2. Intensities were varied from 20–80 dB SPL in steps of 5 dB, with the thickest line corresponding to 80 dB and the thinnest line to 20 dB SPL of the probe tone. Solid lines are used for the probe tone and dashed lines for the suppressor tone (fs). For high fp/fs ratios (e.g., >1.5) the growth of the suppressor response is linear and has a maximum effect in reducing the amplitude of the probe response. As the suppressor tone frequency enters the nonlinear frequency region of the probe (e.g., $fp/fs < 1.46$), the suppressor response exhibits compressive growth when the probe tone intensity is low and linear growth at reduced gain for high probe intensities (thicker dashed lines). For fp/fs ratios ~1 (e.g., fp/fs =1.03), suppression occurs when the amplitudes of the two primaries are approximately equal. When fs is higher than fp

(e.g., fp/fs=0.78), the suppressor tone is in a more compressive region than the probe tone. The probe response is suppressed even though the amplitude of the suppressor's response is considerably less than that of the probe's.

The 2TS-frequency dependence when the probe tone intensity is fixed at 35 dB SPL is shown in Fig. 3(a). As suppressor intensity is increased from 20 to 80 dB SPL [symbols indicate Ls=(the suppressor level in dB SPL)/10] suppression of the probe response is limited to a small frequency range at 30 dB SPL, to over 2 kHz at 35 dB SPL, and continues to increase with increasing Ls in amount and frequency range. At high suppressor intensities the region of suppression encompasses all suppressor frequencies used and clearly would extend to both higher and lower frequencies. Probe amplitude can be reduced by more than a factor of 100 (>40 dB) when $fs < fp$, with the dB reduction being proportional to the intensity difference between the probe and the suppressor ($Ls - Lp$).

If the putative roles of the probe and suppressor are reversed: that is, set the suppressor intensity to 35 dB SPL and then determine the response to the suppressor as fs and Lp are varied, an approximate "mirror image" of the pattern in Fig. 3(a) is seen in Fig. 3(b). That is, for a suppressor frequency (fs) below 6.5 kHz, the rate and amount of suppression decreases with decreasing fs, while for suppressor frequencies above 6.5 kHz there is a faster rate of suppression (ROS) and larger reductions that are largely invariant with frequency. For sufficiently high suppressor frequencies, fs exits from the compressive region of the probe and its

William S. Rhode: Two-tone suppression in chinchilla cochlear mechanics
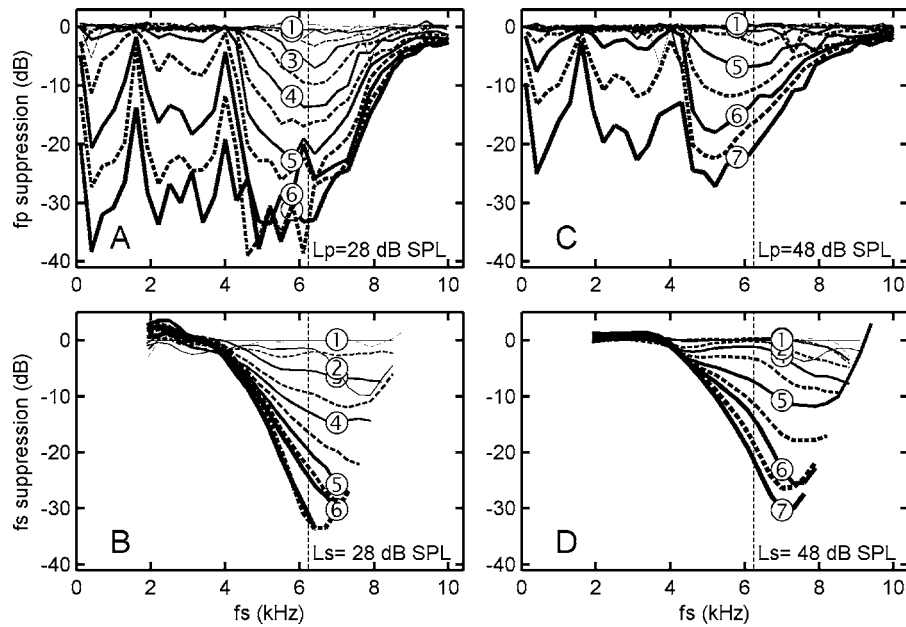
FIG. 4. Mutual suppression at two intensities over a wide frequency range. A. Suppression of responses to a 28 dB SPL probe tone by the suppressor tone as a function of its frequency and intensity. Numeric symbols indicate Ls (in dB SPL/10) rounded to a whole number (e.g., ③=28 dB SPL). Dashed unmarked lines correspond to intervening 5 dB intensities. Curves were normalized by the probe displacement when Ls=20 dB SPL. B. Suppression of the responses to the suppressor tone by the probe tone as a function of suppressor frequency and probe intensities. Numeric symbols as in part A. C. Suppression of responses to a probe tone at 48 dB SPL. D. Suppression of the responses to the suppressor tone by a 48 dB SPL probe tone. Data corresponding to suppressor amplitudes <0.1 nm were considered noise and were eliminated in both C and D. Number of stimulus conditions (fs^*Ls^*Lp): $34 \times 13 \times 13 = 5746$. N143.

response amplitude drops into the noise level and therefore those data were eliminated from the graph. At low probe levels (e.g., Lp=35 dB SPL, dashed line immediately higher than the ④ line), fs is suppressed over the entire 4 kHz range studied at a fairly uniform level of 3–4 dB. There is some suppression (~1 dB) when $fs < fp$ even with the probe intensity 5 dB less than that of the suppressor. The curves indicate that suppression continues to higher frequencies of fs.

The full frequency extent of suppression for a cochlear location with a CF of 6.25 kHz in another animal is shown in Fig. 4 for two probe intensities. Near CF a 28 dB SPL (③) suppressor tone reduces the probe response by ~5 dB over a limited frequency range [Fig. 4(a)]. As the suppressor intensity is increased the region of suppression expands in both frequency directions. With $fp > fs$ there is suppression at high Ls's to nearly an octave above CF. When fs is less than fp, suppression extends to 100 Hz without much variation in amount except for the two notches that are due to the middle-ear transfer function where the same notches occur. The separation between the iso-intensity curves is larger when fs is not in the compressive frequency region (i.e., below 4 kHz). The probe tone reduces the suppressor response as long as $fs > 0.6^* fp$. The probe tone will suppress responses for all $fs > fp$ based on the symmetric relation of the primary tones as mutual suppressors. When the probe intensity is increased 20–48 dB SPL [Fig. 4(d)] the suppression curves are nearly the same as in Fig. 4(b) for Ls at 20 dB lower values. If the suppressive effect was dependent only on the intensity difference, $Lp - Ls$, then it would be expected that these two curves would be very similar.

Rate of suppression (ROS) is a function of the frequency ratio and the relative levels of the primaries and is computed as the slope of the probe-amplitude curves shown in Fig. 2.

ROS curves for high fp/fs ratios show the largest values, approaching 2 dB/dB at high suppression intensities (e.g., for fp/fs=1.46 in Fig. 5). The curves systematically shift to the right as Lp is increased in 5 dB steps from 20 to 80 dB SPL with the thickest curve corresponding to 80 dB SPL. An increase in the maximum ROS occurs when Lp is >60 dB SPL that is then reduced as Ls is increased beyond Lp. ROS amplitudes decrease with decreasing fp/fs ratios with ROS remaining below 0.5 dB/dB for a fp/fs ratio of 0.78. For
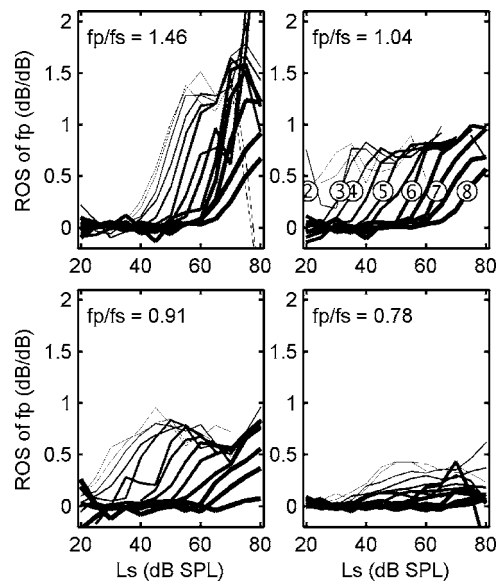
FIG. 5. Rate of suppression (ROS) of the probe tone (at CF=6.2 kHz) is computed as the slope of the probe IO functions for the four suppressor frequencies specified as a frequency ratio fp/fs in Fig. 2. Level of Lp is coded by increasing line thickness. N92.
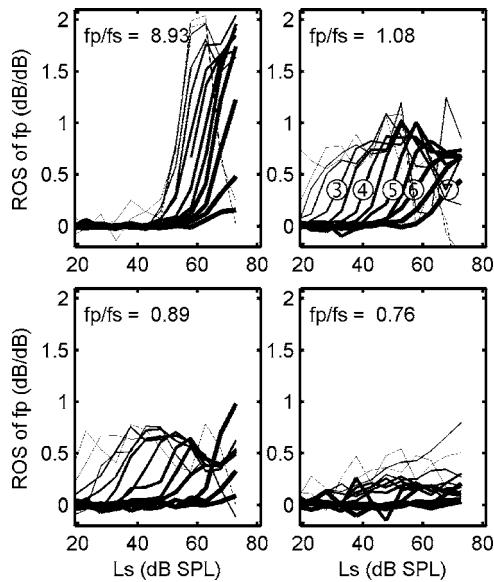
FIG. 6. Rate of suppression (ROS) of the responses to the probe tone (at CF=6.25 kHz), computed as the slope of the probe IO functions for the four suppressor frequencies specified as a frequency ratio fp/fs. Level of Lp is coded by increasing line thickness. N143.

FIG. 7. IO functions for the probe responses (fp) versus intensity of the suppressor (fs). These IO functions are analogous to the suppression curves for auditory nerve fibers used to estimate the suppression rate. The frequency ratios are similar to those used in Fig. 2 (N92). Suppressor intensity ranges from 20 dB SPL for the leftmost curve to 80 dB SPL for the rightmost curve in each panel.

still smaller ratios, ROS continues to decrease and by an octave difference (i.e., $fp/fs=0.5$) there is no measurable suppression effect. A second study shows ROS values of nearly 2 dB/dB for a very low suppressor frequency ($fp/fs=8.93$ in Fig. 6). The maximum ROS values decrease as the fp/fs ratio decreases to ~1 dB/dB close to CF (ratio=1.07) and mostly less than 0.25 dB/dB at a ratio of 0.76.

An alternate calculation of ROS that has been used in the study of suppression in auditory-nerve fibers is the horizontal shift of rate-intensity functions for increasing suppressor levels. Figures 7 and 8 show the horizontal shift of the probe BM-IO function for increasing levels of the suppressor for the frequency ratios indicated in each panel. While there is considerable variability as a function of both the frequency ratio and intensity, it is apparent that there are instances where the horizontal-shift rate of the IO functions as a function of the suppressor intensity can be as large as 2 dB/dB (Fig. 8, ratio=1.86) and that ROS decreases systematically as the fp/fs ratio decreases.

Compatibility between the BM-ROS and AN-ROS is shown in Fig. 9, where a scatter diagram of ROS obtained in 11 animals that had good sensitivity is compared with the standard AN-ROS curve for cat (Delgutte, 1990). The scatter grows as fs/fp, i.e., normalized fs, decreases. Importantly, most of the data points lie above the average curve that Delgutte found for the AN-ROS in cat AN. The scatter covers the same ROS range as that found in cat. The more rapid decrease in BM-ROS with increasing fs could be a species difference (the cochleae are of different lengths). Some of the scatter is also due to the need to average only 2–3 repetitions of the response to a 50 ms stimulus, a necessity when 1–2 h were required to collect the comprehensive suppression data and the limited duration of a high sensitivity cochlear state. Data collected in cochlea with low sensitivity

(i.e., weak compression) had lower values for ROS across the entire intensity-frequency region and were not included in Fig. 9.

C. Suppressor amplitude for 1 dB suppression of the probe

One difference between AN and BM suppression at low frequencies has been stated to be that the intensity of the BM suppressor appears to always be higher than that observed in comparable AN frequency threshold curves (Cooper, 1996). This is not the case in four sensitive chinchilla cochleae (Fig.
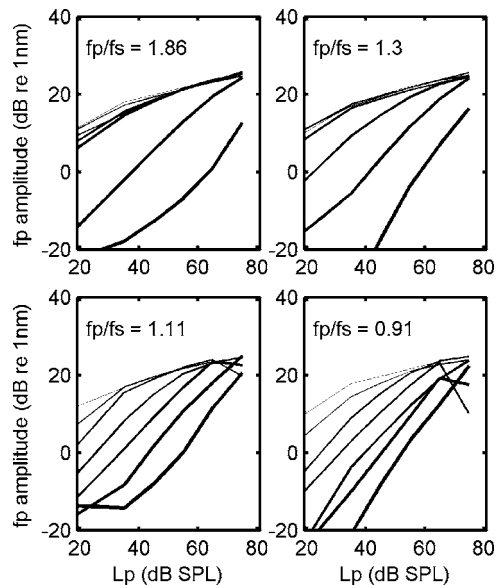
FIG. 8. IO functions for the probe responses (fp) versus intensity of the suppressor (fs) for another animal (N157). Intensities were 15–75 dB SPL in 10 dB steps. CF=9.1 kHz.
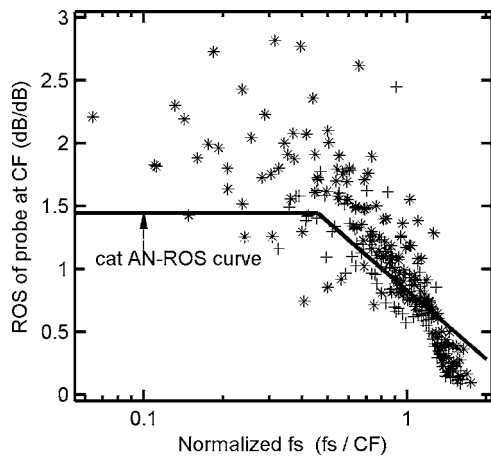
FIG. 9. Scatter diagram of the maximum rate of suppression (ROS) for 11/animals. The solid line is the Delgutte fit to ROS auditory-nerve data (Delgutte, 1990). Asterisks (*, $n=7$) indicate data from high-sensitivity cochleae (gain >55 dB) and plus symbols (+, $n=4$) indicate data from slightly lower sensitivity cochleae with gains between 40 and 50 dB. Individual data points obtained from ROS functions with the probe at CF and 20 dB SPL. In three experiments the probe intensity was 30 dB SPL due to the absence of 20 dB data.

10). At the same tonal intensities, suppressor response amplitudes needed to produce 1 dB of probe response reduction (dashed lines) were usually less than the probe amplitude (solid lines) across the entire frequency range. The threshold amplitude of the suppressor is nearly constant when the suppressor frequency is less than CF/2.

D. Relation of ROS to ROG and BM tuning

Iso-intensity functions for pure tones at 60 and 70 dB SPL are indicated by dashed lines in Fig. 11(a) (right ordinate). Superimposed are ROS functions when the intensity of the probe (at CF) is as indicated by the numeric symbols
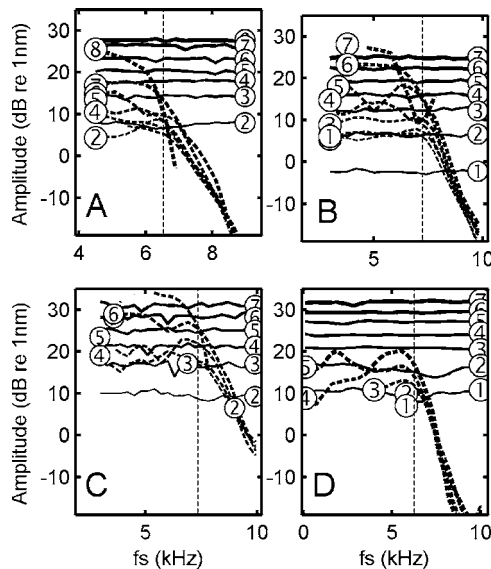
FIG. 10. Displacement amplitude functions for both the probe (solid lines) and suppressor (dashed lines) as a function of the suppressor frequency and intensity for a 1 dB reduction in probe amplitude. Intensities/10 for probe and suppressor, respectively, are indicated by the numeric symbols. A. N92. B. N96. C. N118. D. N143.
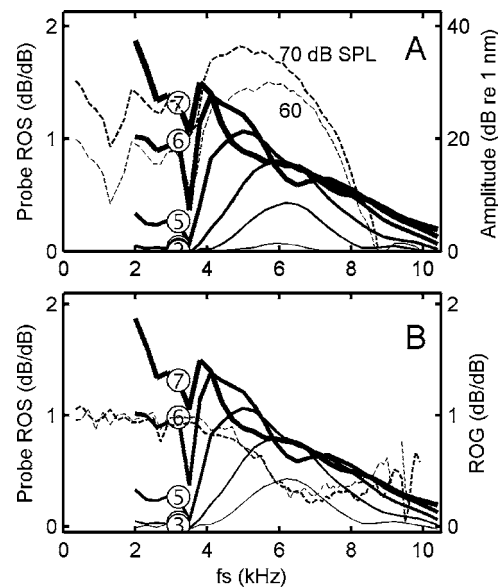
FIG. 11. A. Comparison of the rate of suppression (solid lines) to two sample iso-intensity displacement curves (dashed lines). The former are shown for 20–80 dB SPL in 20 dB steps indicated by the usual numeric symbols. B. ROS (solid lines) is compared to ROG (dashed lines). CF =6 kHz. N136.

($Lp=$(dB SPL)/10). ROS increases as fs decreases and is influenced by the iso-intensity dips and peaks that are largely due to middle-ear amplitude variations. ROS (solid lines) and rate-of-growth (ROG) (dashed lines and right ordinate) functions are superimposed in Fig. 11(b). For low-frequency suppressors and intensities >60 dB SPL, ROG is 1 dB/dB and ROS is >1 dB/dB. Suppression continues for fs's higher than the response region for single tone excitation. There is a similar pattern in the two functions up to 8 kHz, a region where high single-tone intensities can produce notches in the IO functions.

E. Phase of maximum suppression

Figure 12(a) illustrates the phase-frequency relation when the probe tone is at the specified intensity in each panel and the suppressor intensity is varied. With $Lp=33$ dB SPL, there is a phase lag over nearly the entire frequency range with maximum values of ~0.5 cycle (π radians). Probe phase begins to lag when $Ls=(Lp+15$ dB SPL) except in the vicinity of 8 kHz where a lag appears to be present at low suppressor intensities. In the comparable plot for the suppressor response [Fig. 12(b)], where the roles of suppressor and probe are reversed, there are increasing suppressor phase lags up to ~6.5 kHz. For higher suppressor frequencies there is an increasing phase lead with increasing probe intensity. These curves are analogous to single-tone phase functions where one expects increasing lags below CF and increasing phase leads above CF with increasing intensity. Note that the sharp phase changes at ~2 and 4 kHz are due to the middle-ear phase function.

In contrast to phase behavior in very sensitive cochleae, the probe phase sports a small phase lead (<25°) with increasing Ls in less sensitive cochleae [e.g., Fig. 12(c) and 12(d)]. However, for $Ls>60$ dB SPL probe phase lags for
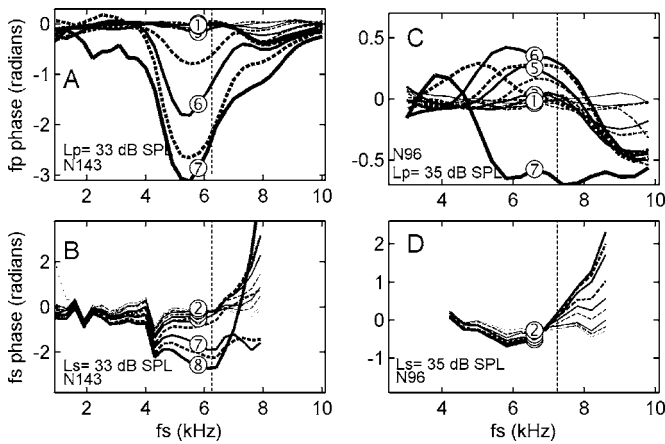
FIG. 12. Normalized probe response phase as a function of the suppressor frequency and intensity. A. Probe is at 33 dB SPL and suppressor intensity is indicated by the numeric symbols. (e.g., ⑥=(63 dB SPL)/ten rounded down). B. Suppressor response phase when its intensity is 33 dB SPL and intensity of the probe is indicated by the same symbols. A and B. N143 had a high gain, 58 dB. C and D same as in A and B for N96 which had a lower gain=30 dB. For Fig. 12(c) the level of the suppressor tone is indicated by the usual symbols (Ls in dB SPL/10, rounded) and varies between 20 and 80 dB SPL.

nearly all fs. The suppressor phase functions are similar to those of high sensitivity cochleae though the phase change is smaller.

Instantaneous phase effects can be studied by separating the suppressor and probe wave forms using Butterworth filters that are chosen so as to not alter individual amplitudes and correcting for any phase distortions introduced by the filters. The effects of suppressor tones at four frequencies and three intensities on the probe response wave form are illustrated in Fig. 13. The bottom panel of the first column illustrates the wave form for a 62 dB SPL suppressor tone of 100 Hz and 40 dB SPL probe tone (Ls:Lp=40:62 dB SPL). Maximum suppression occurs when the BM is displaced to-

ward scala vestibuli (SV). When suppressor intensity is increased to 72 dB SPL, there is nearly a symmetric reduction in amplitude of the probe at maximum displacement of the 100 Hz suppressor. When Lp:Ls=80:77 dB SPL, suppression maxima occur for displacement maxima toward either scala and there is suppression throughout the period of the suppressor. With a suppressor frequency of 200 Hz and Ls:Lp=70:81 dB SPL, maximum suppression is for displacement maxima toward either cochlear scala with a small phase lag (col. 2). With increasing suppressor intensities the pattern remains the same. With a suppressor frequency of 700 Hz (col. 3) and Lp:Ls=55:65 dB SPL, the maximum suppression occurs with a displacement toward SV. Increasing Ls by 10 dB results in a sustained amplitude reduction of the probe tone and the maximum suppression remains at the maximal BM displacement toward the SV. When suppressor intensity was increased to 90 dB SPL there was nearly complete suppression of the probe with equal amounts of suppression for displacements toward either scala. There is a similar pattern for a 1000 Hz suppressor (col. 4), though for the highest suppressor intensity, 90 dB SPL, maximum suppression occurs for a displacement toward the scala tympani and the probe wave form is very distorted. A problem with determination of the phase of maximum distortion is that suppression is not instantaneous, rather there appears to be a time lag following SV displacement before the maximum effect appears in the probe signal.

There is a transient in BM vibration when the probe tone is turned off (with a 1 ms fall time) that lasts for at least 9 ms [Fig. 14(a)]. When a 200 Hz suppressor is also applied that lasts 10 ms longer than the probe tone, it not only suppresses somewhat the continuous portion of the probe tone's response but largely terminates ringing at the probe tone frequency [Fig. 14(b)].
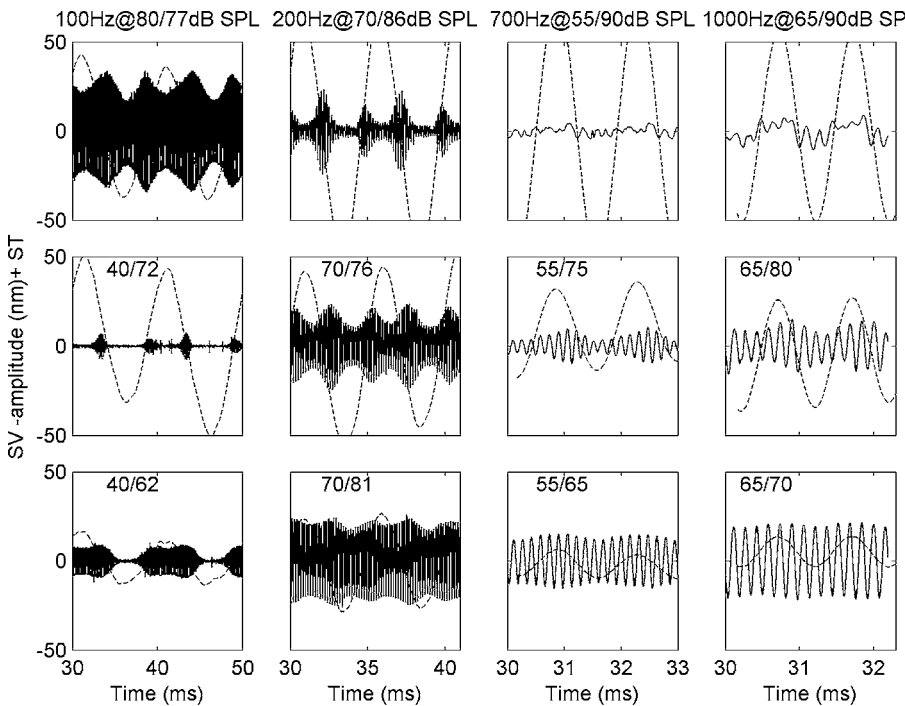
FIG. 13. Response wave form components during presentation of four different suppressor tones, at the frequency indicated above each column. Intensities are given for each panel as (Ls, intensity of suppressor tone):(Lp, intensity of probe tone at CF =6350 Hz). The response to the probe was obtained by applying a fourth order high-pass Butterworth filter. The response to the suppressor tone was obtained by applying a third order band-pass Butterworth filter. Corner frequencies were chosen so as to not attenuate either component. The phase of the wave form was corrected for the phase of the filter. Solid line is the CF component and the dashed line is the suppressor component. The suppressor wave form is shown at half amplitude in the top row. N136.
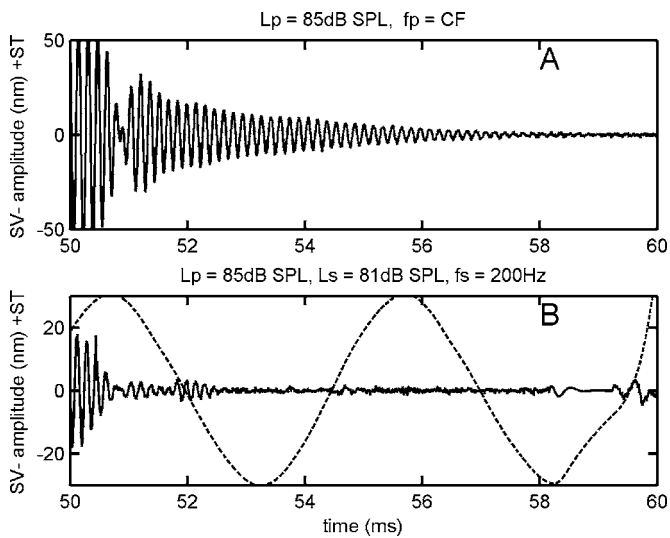
FIG. 14. A. Illustration of the off response for an 85 dB SPL tone 50 ms in duration. B. The effect of a 200 Hz suppressor tone on the response to the 85 dB SPL CF probe tone when the suppressor tone was at 81 dB SPL. CF=6 kHz, N136.

F. Iso-suppression functions

Suppression frequency functions of the probe response for a constant 1 dB reduction are compared to 1 and 4 nm iso-amplitude single-tone functions for four separate studies in Fig. 15. The suppression functions are much wider in frequency than the single-tone BM iso-amplitude filter functions. However, at low probe levels and $fs < CF$, the two functions have approximately the same slope, as best shown in Figs. 15(b) and 15(d). With increasing probe intensity, e.g. $Lp > 40$ dB SPL, the iso-suppression function is relatively
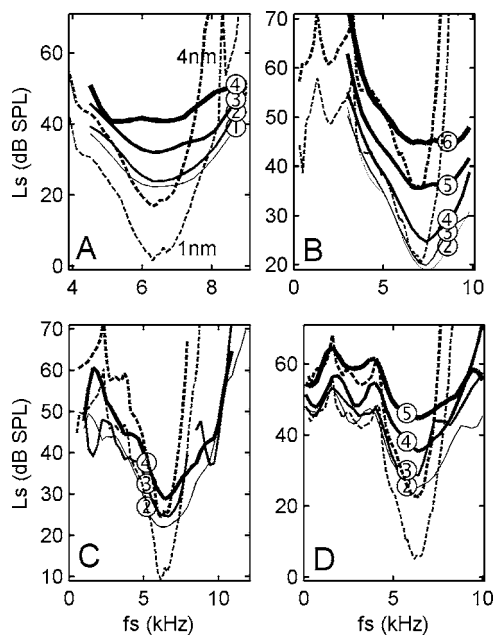
FIG. 15. Suppressor intensity required to effect a 1 dB reduction in probe amplitude for four probe intensities (solid lines, where the symbols indicate Lp in dB SPL/10). Superimposed are single-tone iso-displacement curves for a 1 and 4 nm amplitude indicated by the dashed lines. Thicker line =4 nm. A. N92. B. N96. C. N121. D. N143.
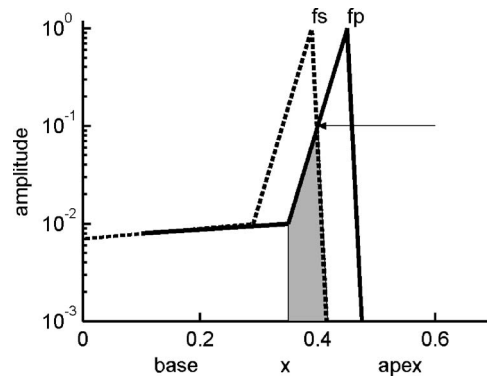
FIG. 16. The spatial amplitude distribution for the probe and suppressor tone thresholds along the basilar membrane ($fs > fp$). The tail region of the distributions is shown as nearly flat as a function of location because the threshold curves were not normalized with respect to the middle-ear displacement.

flat in the CF region. This flat curve suggests that the probe filter function is not the sole determinant of the region of two-tone interaction.

When fs is in the low-frequency (tail) region of the BM filter, the suppression curve has nearly the same shape as the iso-amplitude curve [especially Figs. 15(c) and 15(d)]. Often the suppressor intensities were not taken at low enough intensities for a valid comparison of the iso-suppression function to be made with the 1 nm iso-amplitude function. When sufficiently low intensities (i.e., 20 dB SPL) were used, as in Fig. 15(b), iso-suppression and iso-amplitude curves are similar for $fs < fp$. Above CF there is a lower slope to the suppression data and the suppression curve extends over a much wider frequency range than the iso-amplitude curve.

IV. DISCUSSION

A. Goals of this study

The principal intents of this study were to extend previous observations of 2TS in the mammalian cochlea to the mid-frequency (~6 kHz) region of sensitive cochleae and to address issues of 2TS that were unresolved in previous mechanical and neural studies. Most of the earlier observations of BM-2TS were made in the base or hook region of the cochlea, a very high frequency region, where it is usually very difficult to maintain high cochlear sensitivity during the experiment. As a result, 2TS in this region was often shown to produce modest suppression, often less than 20 dB, an amount considerably less than found in AN-2TS studies.

Results of the present study:

1. Mutual suppression. Sensitive cochlea have high gains and strong compression with single-tone ROGs as small as 0.1 dB/dB at CF. This study [e.g., Fig. 3(a)] shows that suppression of the probe response in sensitive cochleae can exceed 40 dB for all suppressor frequencies (fs) less than the probe-tone frequency (fp=CF). Suppression magnitude drops off as fs increases above CF at a rate that corresponds to the low-frequency slope of the tip of the BM threshold iso-amplitude curve (equivalent to a mechanical frequency threshold curve if displacement is the effective stimulus for an ANF). Figure 16 illustrates idealized BM

threshold amplitude curves along the cochleae for both the probe and a higher frequency suppressor. The threshold amplitude of the suppressor response as a function of increasing fs decreases at the same rate as the low-frequency leg of the tip region of the probe threshold curve. That is, the rates are 65–75 dB/oct for the four plots in Fig. 10. This corresponds closely to the curve traced out by the intersection of the dashed and solid lines in Fig. 16 as fs in increased ($fs > fp$). It is a product of the overlap of the spatial distribution of fs with that of the nonlinear region of fp that is indicated by the shaded region in Fig. 16.

Suppression extends from the lowest suppressor frequencies used to nearly an octave above CF. That is, as long as there is displacement of the BM due to the suppressor in the region where the probe IO functions are compressive, the probe response will be suppressed. Suppression extends throughout and beyond the single-tone tuning region, a fact that was anticipated by AN two-tone studies (Hind *et al.*, 1967). Mutual suppression has been also been shown in intracellular recording from inner hair cells (IHCs) (Sellick and Russell, 1979).

Either primary tone can "suppress" the other's response, hence there is mutual suppression, that extends over more than an octave region of the cochlea with the appropriate stimulus parameters. When $fp > fs$, the probe tone cannot suppress the response to a low-side suppressor unless the response to the probe overlaps the nonlinear response region of the suppressor (Fig. 16). The effect of the probe on the response to a suppressor is nearly a mirror image of the suppressor's effect on the response to the probe (e.g., Fig. 3). From the peak of mechanical tuning for a probe tone, the basalward extent that a higher frequency tone can suppress the response the probe tone is approximately 1.8 mm in the 6 kHz frequency region. For reference, an octave in the 6 kHz region spans 3.4 mm (Greenwood, 1990). In contrast, a low-frequency tone of sufficiently high intensity will suppress the response to any high-frequency probe tone.

2. Suppression threshold. At the threshold of suppression, the intensity of the suppressor when its frequency is near CF can be less than the intensity of the probe tone being suppressed (Figs. 3 and 4). Similarly, the intensity of a suppressor necessary for a 1 dB reduction of the probe response in the tail region of the BM tuning curve can be at or below the intensity necessary to obtain the nominal 1 nm threshold displacement of the BM at the tail frequencies [Figs. 15(c) and 15(d)]. As the threshold criterion for suppression of the probe response is raised, e.g., to 4 nm, the suppressor intensity necessary to reduce the response to the probe by 1 dB is often less than the single-tone intensity in the tail region required to produce the criterion displacement. In contrast, in the high-frequency region of the guinea pig cochlea the intensity of the suppressor at suppression threshold in the tail frequency region was nearly 20 dB above the single-tone tail frequency excitatory threshold (Cooper, 1996). In the present study, the suppressor amplitude threshold for a 1 dB reduction of a CF tone with 1 nm amplitude is on the order of 2 nm which agrees with a 2 nm threshold in guinea pig from Geisler and Nuttall (1997) and with the 1–5 nm range of Cooper (1996). If there was a loss of sensitivity at CF, as is

often observed in AN fibers that have higher than normal thresholds, then the threshold suppressor intensity increases also, though it can be less than the corresponding 'threshold' CF intensity [Fig. 15(b)]. The shape of the suppressor threshold functions in the tail region follows that of the single tone iso-intensity function [Figs. 15(c) and 15(d)] in agreement with previous observations on gerbil AN fibers that showed low-side suppression threshold was nearly constant as a function of suppressor tone frequency (Schmiedt, 1982). The first indication of suppression, a small phasic reduction of the probe response, occurs when the suppressor response amplitude is less than that of the probe (e.g., lower right panels in Fig. 13). In the high-frequency region of the guinea pig cochlea, the threshold suppressor response amplitude was always equal to or greater than that of the probe (Geisler and Nuttall, 1997).

3. Rate of suppression. The magnitude of rate-of-suppression (ROS) and the range of values are largely in accord with the ROS found in the auditory nerve of cat (Delgutte, 1990; Pang and Guinan, 1997). ROS's are on average somewhat larger than found in cat for $fs < CF$ and smaller when $fs > CF$ (Fig. 9). The latter feature could be due to the difference in the lengths of the two cochleae, where fewer hair cells appear to participate in the nonlinear region in chinchilla than in cat. The higher low-frequency average ROS in chinchillas could also be a result of selecting the results from only very sensitive cochleae for the present report. Some of the variability in the data presented here is undoubtedly due to the use of only 2–3 repetitions of short tones resulting in higher variability than when ten repetitions are used, as in single-tone responses. The use of few repetitions was necessary as the duration of the results of a 2TS paradigm, even with 2–3 repetitions per stimulus pair, was often of the order of 2 h. Any noise in the measurements often results in large increases in ROS variability when the gradient (= derivative) of the IO functions is calculated. Therefore, obvious "wild points" in the gradient functions were deleted. The systematic decrease in ROS when $fs > CF$ differs from an earlier study that suggested that there is a constant ROS of around 0.5 dB/dB for fs's $> CF$ (Costalupes *et al.*, 1987). This latter study also found that ROS was 1.4 dB/dB for low suppressor frequencies, a value very near the average found for cat AN (Delgutte, 1990; Pang and Guinan, 1997). For low-frequency suppressors (with ROG of 1 dB/dB) suppressing a probe tone that has a ROG between 0.1 and 0.4 dB/dB, one might conjecture that ROS's should be between 2.5 and 10 dB/dB [i.e., 1/(probe ROG)], higher values than typically obtained in any study. Another explanation for the lower than estimated ROS's is a result of mutual suppression, whereby peak ROS values are reduced due to linearization of the response to both tones. Reduced ROS could result from lower than optimum sensitivity of the preparation, which is often not reported. In the worst case, ROS is zero in a linear cochlea (Ruggero *et al.*, 1992). It is concluded that large ROS's will be obtained only in 2TS studies conducted in highly sensitive, uncompromised cochleae.

4. Phase shift. The effect of an increase in intensity of the suppressor or probe on the phase of the probe or suppres-

sor, respectively, is an increasing phase lag as large as 180° in eight high sensitivity cochleae [e.g., Fig. 12(a)]. In four cochleae with lower gain there was a "small" phase lead with increasing suppressor intensity [e.g., Fig. 12(c)]. This latter result is similar to that found in the guinea pig (Cooper, 1996). The maximum phase change, $\sim \pi$ phase lag, observed in the present study [Fig. 12(a)] is larger than found previously and is at odds with earlier studies. The underlying reason is not known.

5. Phase dynamics. The effect of suppression intensity on probe phase for very low-frequency suppressors shows that maximum BM displacement results in maximum suppression similar to several previous studies (Cooper, 1996; Geisler and Nuttall, 1997). However, while there is a consensus that the maximum suppression occurs for displacements of the BM toward the scala tympani, the present results indicate that there are often nearly equal amounts of suppression associated with BM displacements toward either scala. Direction preferences vary with probe intensity and suppressor frequency and even for a single suppressor frequency. When the amplitude of the suppressor response is equal to or less than that of the probe, usually the first sign of suppression of the probe response, i.e., amplitude modulation of the probe response, occurs when the BM is displaced towards SV. Irrespective of the suppressor frequency, for suppressor response amplitudes greater than ~ 25 nm there was a tonic reduction of the probe response. Note that there are two peaks in the envelope of the probe response at the suppressor displacement minima (maximum velocity) of the response to the suppressor (e.g., Fig. 13, 100 Hz and $Lp:Ls=40:40$) The double peak also occurs in cycle histograms in the response of AN fibers to 2TS stimuli (Temchin *et al.* 1997).

B. Low-frequency suppressive effects

There remains a difference between mechanical and neural two-tone rate suppression for suppressors low in frequency relative to the probe tone at CF. It has been noted that the maximum displacement of the basilar membrane for this condition is never smaller than the maximum of the unsuppressed response amplitude (Cooper, 1996; Geisler and Nuttall, 1997). This leaves undecided the mechanical correlate (if any) for rate suppression in the auditory nerve. Low-frequency suppressors can increase the threshold for an AN response to a CF tone by 50–70 dB (Delgutte, 1990; Javel, 1981). In contrast, there is no evidence to date that maximum BM vibration amplitude for the comparable experiment can be reduced even 1 dB for the combined low-frequency suppressor and CF wave form. Cheatham and Dallos (1999) provided a possible mechanism by which a low-frequency suppressor could be reduced by showing that the inner hair cells are driven by velocity of the fluid displacement rather than by displacement at frequencies less than 200 Hz. A fluid velocity drive is supported by indications that the cilia of the IHCs are not embedded in the tectorial membrane as the cilia of the OHCs are and therefore would likely not be driven by BM displacements (Slepecky,1996). The conclusion reached by several investigators is that there has to be some sort of a second filter to account for the elimination or reduction of

the responses to the low-frequency suppressor (Cooper, 1996; Geisler and Nuttall, 1997; Sen and Allen, 2006; Robles and Ruggero, 2001). A velocity drive would be equivalent to a high-pass filter and would attenuate the low-frequency suppressor. As an alternative, Hill (1998) argued that a nonmechanical factor that is derived from the suppressor tone must exert a suppressive influence on the auditory nerve-fiber afferents. He proposed dc current spread due to the suppressor could affect spike generation. Another possibility is that the summating potential (SP) affects the transmembrane potential thereby biasing the synaptic events. SP is a dc potential that likely arises from both OHCs and IHCs due to rectifying properties of each (Durrant *et al.* 1998). Yet another possibility is that there is an effect of the fluid flow in the tunnel of Corti (Karavitaki and Mountain, 2003). Vibration of the bony spiral limbus could also affect filtering of BM displacement before the IHCs are stimulated (Rhode, 2007). The amplitude of the BM/limbus transfer ratio decreases as *fs* decreases at an average rate of ~ 3 dB/oct (Rhode, 2007). At some low *fs* (<500 Hz) the ratio will approach zero and result in a significant reduction of excitation of the hair cells.

It has been reported that most low-frequency mechanical suppression thresholds were 10–20 dB above the threshold iso-displacement curves in the basal region of the guinea pig cochlea (Cooper, 1996). On the other hand, AN-2TS studies in chinchilla conclude that low-side suppression corresponds to a relatively constant BM amplitude movement and that modulation of the probe response and rate suppression occur at nearly identical intensities irrespective of spontaneous rate (Temchin *et al.* 1997). Pang and Guinan (1997) did not find any rate-suppression dependency on spontaneous rate of AN fibers in cat. In contrast, Cai and Geisler (1996) found that threshold suppressor intensity for low-spontaneous AN fibers was substantially below that of the excitatory threshold and that the suppression threshold was relatively constant with frequency of the suppressor. Any suppression dependency on the spontaneous rate of the AN fibers may simply the explained by the fact that low spontaneously active AN fibers have higher thresholds than high-spontaneous fibers (Liberman, 1978). However, their result is controversial.

Examination of the entire set of wave forms ($n=1690$) corresponding to Fig. 13, demonstrated that BM suppression threshold is relative to the probe's response amplitude. For example, for a 90 dB SPL probe at 300 Hz, the probe and combined-wave form suppression threshold amplitudes are 50 and 90 nm, respectively. In comparison, for a 30 dB SPL probe they are 5 and 7 nm, respectively. The tail region of the BM frequency threshold curve is relatively flat and the suppression response threshold is also nearly flat [cf. Fig. 15(d)]. This result is in agreement with AN-frequency-threshold curves which do not vary much in the tail region and as a result suppression thresholds in the tail region are also nearly flat (Schmiedt, 1982).

For low-frequency suppressors, it is clear that, at suppressor/probe intensity ratios of >40 dB, the probe response wave forms were no longer simple sinusoids (Fig. 13). Geisler and Nuttall (1997) found as the suppressor frequency increased that the largest amplitude associated with the suppressor was at the lowest fundamental frequency of

the suppressor and probe frequencies. They hypothesized that the cochlear amplifier may not be fast enough to follow high suppressor frequencies. In the present study, Fourier analysis of the resultant wave forms when the suppressor response amplitude is large (cf. Fig. 13) shows that there is a large amount of distortion introduced including harmonics of the suppressor frequency but also combination tones at $fp \pm n^*fs$ (not shown). This likely occurs due to rectifying characteristics of the OHC-IO curve and the resulting feedback into BM motion.

A relatively simple "saturation feedback model" that incorporated the nonlinear compressive IO curve of OHCs resulted in fairly realistic results for two-tone suppression (e.g., Geisler and Nuttall, 1997). The operating point of the OHCs becomes a key factor in determining whether suppression occurs initially with BM motion towards the SV or scala tympani. In the present results it appears that the operating point was fairly near the center of a nearly anti-symmetric OHC-IO curve.

C. High-frequency suppressive effects

For suppressor tones with frequencies greater than CF, their responses at the CF location can be very small relative to that of the probe (<-40 dB). This implies that the principal site of two-tone suppression lies basal to the probe's CF location (Geisler *et al.*, 1990; Cooper, 1996). Note that the amplitude of the suppressor's response at its more basal CF location is considerably larger than at the probe location. Given the rate of decrease in the amplitude of the suppressor, it can be concluded that as the overlap of the suppressor envelope with that of the probe decreases, the amount of suppression decreases. The fact that there can be suppression when the suppressor frequency is nearly an octave above the probe frequency suggests that there remains some residual low amplitude of displacement due to the suppressor ($fs > fp$) within the compressive region of the probe tone. This idea results from single tone studies where the ROG for frequencies higher than CF is near 0 dB/dB allowing the possibility that these frequencies move the BM for considerable distance apical to their CF location though with very small amplitudes.

With increasing intensity the plateau effect becomes apparent in BM responses to pure tones, but may not play a role in the suppression of the probe as it likely results from another form of excitation, either a fast compressive fluid-conducted wave (Lighthill, 1981; Andoh and Wada, 2004) or an evanescent wave defined as one "that decays exponentially with distance and has constant phase with increasing frequency and therefore a very high velocity" (Raichel, 2000). A high-frequency plateau has not been seen in AN tuning curves but has been reported in intracellular OHC recordings (Kössl and Russell, 1992). Because the limbus vibrates at plateau frequencies at amplitudes usually 10 dB less than the BM, it probably is not sufficient to account for the AN results. Therefore, it remains likely that there is additional processing inserted somewhere between the IHCs and the BM-OHC complex to remove the plateau response from driving IHCs.

D. Rate of suppression for low-frequency suppressors

There was a considerable discrepancy between BM-ROS and AN-ROS for low-frequency suppressors in the literature. Ruggero *et al.* (1992) reported that 1.4 dB/dB was the norm for BM suppression. Sen and Allen (2006) suggested that a comparison of the latter value with values of 2.4 dB/dB and higher found in AN-2TS studies may require factors other than a simple basilar membrane origin of neural 2TS. However, the average ROS that Delgutte (1990, Fig. 13) found for low-frequency suppressors was ~1.4 dB/dB with a range of 0.2–2.8 dB/dB. In comparison, Rhode and Recio (2001) showed a 2 dB/dB ROS for a CF region of 8 kHz. Present results (Fig. 9) show low-frequency ROSs at least up to 2.5 dB/dB in the 6 kHz cochlear region. In fact, the average value obtained for the BM-ROS in sensitive chinchilla cochleae appears larger than found in one cat study AN-ROS (Fig. 9) and similar to the results of a study that used a low-frequency noiseband as the masker (Pang and Guinan, 1997). Further, Delgutte showed that ROS for low-frequency AN suppressors increased with increasing CF implying the ROS found for BM for the 6 kHz region likely underestimates the maximum ROS which can exist in the cochlea. The conclusion is that the ROS produced by low-frequency suppressors found for BM-2TS vibration is entirely adequate to explain the comparable ROS in the AN without the introduction of additional mechanisms.

E. Amplitude of suppression

Several publications of BM-2TS reported that the maximum suppression was in the range of 10–30 dB while auditory nerve studies have shown as much as 70 dB using a synchrony suppression measure (Javel, 1981; Pang and Guinan, 1997). In the present study, suppression >40 dB occurred [cf. Fig. 3(a)] and it is likely that suppression continues to much larger values. It may be difficult, if not impossible, to detect these larger amounts of suppression with the present techniques unless many more responses are averaged than were done here. This problem results from the large range of the signals (~100 dB) and the limited range of the 16 bit A/D converter used to capture the signals. A low level probe signal is often encoded using only a few bits of the converter, so with suppression its amplitude can be smaller than that corresponding to a single bit of the converter. Significant averaging would be required to pull out a statistically significant result. For example, if the suppressor amplitude was to be as much as 60 dB greater than that of the probe, the probe signal would be represented in the lower 5 bits without suppression. With 60 dB of suppression the maximum probe amplitude would be ~1/30 of the least significant bit and even with 40 dB of suppression the probe signal maximum would be only a fraction of the least significant bit. In the present study, the goal was to cover a large range of frequencies and amplitudes so that only 2–3 responses were averaged preventing tracking suppression for very large ratios of suppressor-to-probe amplitude. However, in a two-tone suppression experiment, Geisler and Nuttall

(1997) averaged signals for 20 s and obtained suppression values as large as 60 dB, an amount that is compatible with suppression values in AN.

These combined results suggest that BM-2TS behavior is sufficient to explain any AN-2TS suppression as long as the frequency of the suppressor is not in the tail region of the tuning curves. Based on the number of possible explanations, the jury is still out on how low-frequency rate suppression in AN fibers takes place.

F. Threshold of suppression

Another 2TS feature that was previously in doubt resulted from estimates of the amplitude of displacement due to the suppressor that produced just detectable suppression of the probe response. Cooper's (1996) study of suppression in the basal region of the guinea pig cochlea indicated that the BM required a mechanical displacement produced by the suppressor that was 10–20 dB more that expected based on AN studies. However, the data were collected mostly in preparations that were in less than optimal condition. By focusing on very sensitive cochleae as this study did, BM and AN response properties are in much better agreement. This does not even take into consideration what additional closure might be possible if the same AN-2TS studies were performed in the same preparation as the mechanical studies, such as done for single-tone stimuli by Narayan *et al.* (1998).

ACKNOWLEDGMENTS

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects," J. Acoust. Soc. Am. **49**, 1131–1139.

Andoh, M., and Wada, H. (**2004**). "Prediction of the characteristics of two types of pressure waves in the cochlea: Theoretical considerations," J. Acoust. Soc. Am. **116**, 417–425.

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

Carr, C. D., and Bohne, B. A. (**1979**). "Location of structurally similar areas in chinchilla cochleas of different length," J. Acoust. Soc. Am. **66**, 411–414.

Chan, D. K., and Hudspeth, A. J. (**2005**). "Ca^{2+} current-driven nonlinear amplification by mammalian cochlea *in vitro*," Nat. Neurosci. **8**, 149–153.

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

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

Cooper, N. P., and Rhode, W. S. (**1992**). "Basilar membrane mechanics in the hook region of cat and guinea pig cochlea: Sharp tuning and nonlinearity in the absence of baseline shifts," Hear. Res. **63**, 163–190.

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

Cooper, N. P. (**1999**). "Vibration of beads placed on the basilar membrane in the basal turn of the cochlea," J. Acoust. Soc. Am. **106**, L59–L64.

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

Delgutte, B. (**1990**). "Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Durrant, J. D., Wang, J., Ding, D. L., and Salvi, R. J. (**1998**). "Are inner hair cells the source of summating potentials recorded from the round window?," J. Acoust. Soc. Am. **104**, 370–377.

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

Fay, R. R. (**1988**). *Hearing in Vertebrates: a Psychophysics Databook* (Hill-Fay, Winnetka, IL).

Geisler, C. D., Yates, G. K., Patuzzi, R. B., and Johnstone, B. M. (**1990**). "Saturation of outer hair cell receptor currents causes two-tone suppression," Hear. Res. **44**, 241–256.

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

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hill, K. G. (**1998**). "Basilar membrane motion in relation to two-tone suppression," Hear. Res. **115**, 129–142.

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

Hudspeth, A. J., and Jacobs, R. (**1979**). "Stereocilia mediate transduction in vertebrate hair cell," Proc. Natl. Acad. Sci. U.S.A. **76**, 1506–1509.

Javel, E. (**1981**). "Suppression of auditory nerve responses I: Temporal anlaysis, intensity effects and suppression contours," J. Acoust. Soc. Am. **69**, 1735–1745.

Karavitaki, K. D., and Mountain, D. C. (**2003**). "Is the cochlear amplifier a fluid pump?," In: *Biophysics of the Cochlea. From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 310–311.

Kössl, M., and Russell, I. J. (**1992**). "The phase and magnitude of hair cell receptor potentials and frequency tuning in the guinea pig cochlea," J. Neurosci. **12**, 1575–1586.

Liberman, M. C. (**1978**). "Auditory nerve responses from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **83**, 93–98.

Lighthill, J. (**1981**). "Energy flow in the cochlea," J. Fluid Mech. **106**, 149–213.

Lin, T., and Guinan, J. J., Jr. (**2004**). "Time-frequency analysis of auditory-nerve fiber and basilar membrane click responses reveal glide irregularities and non-characteristic-frequency skirts," J. Acoust. Soc. Am. **116**, 405–416.

Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A. (**1998**). "Frequency tuning of the basilar membrane and auditory nerve fibers in the same cochleae," Science **282**, 1882–1884.

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

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

Raichel, D. R. (**2000**). *The Science and Application of Acoustics* (AIP, Springer-Verlag, New York).

Rhode, W. S., and Cooper, N. P. (**1993**). "Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae," Hear. Res. **66**, 31–45.

Rhode, W. S., and Recio, A. (**2000**). "Study of the mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.

Rhode, W. S., and Recio, A. (**2001**). "Multicomponent stimulus interactions observed in basilar membrane vibration in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **110**, 3140–3154.

Rhode, W. S. (**2007**). "Basilar membrane mechanics in the 6–9 kHz region of sensitive chinchilla cochleae," J. Acoust. Soc. Am. **121**, 2792–2804.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Ruggero, M. A., Robles, L., and Rich, N. C. (**1992**). "Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression," J. Neurophysiol. **68**, 1087–1099.

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

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

Sen, D., and Allen, J. B. (**2006**). "Functionality of cochlear micromechanics—as elucidated by upward spread of masking and two tone suppression," Acoust. Aust. **34**, 37–42.

Slepecky, N. B. (**1996**). "Structure of the mammalian cochlea," In: *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 44–129.

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

Zheng, J., Shen, W., He, D. Z. Z., Long, K. B., Madison, L. D., and Dallos, P. (**2000**). "Prestin is the motor protein of cochlear outer hair cells," Nature (London) **405**, 149–155.

Zwicker, E. (**1979**). "A model describing nonlinearities in hearing by active process with saturation at 40 dB," Biol. Cybern. **35**, 243–250.

Forward and reverse waves in nonclassical models of the cochlea

Egbert de Boer[a]

Room D2-226, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam,
The Netherlands

In "classical" models of the cochlea the mechanical properties of the cochlear partition (in which the basilar membrane plays a dominant role) depend only on the local longitudinal coordinate. Wave propagation in such models is the same for both propagation directions, towards the apex and towards the base. "Nonclassical" models contain mechanical properties that depend on variables at more than one location. In such models the properties of wave propagation may well be asymmetrical. This note presents an analysis of feed-forward and feed-backward models, with the emphasis on the properties of forward and reverse waves. In both of these classes of model, wave amplification for forward waves turns into wave attenuation for reverse waves. The slope of the phase pattern is nearly the same for the two types of waves, it is increased in feed-forward models and decreased in feed-backward models—with respect to a model without feedback.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2715743]

I. INTRODUCTION

A "classical" model of the cochlea is entirely controlled by the *local* basilar-membrane (BM) impedance, to be designated by $Z_{BM}(x, \omega)$ where x is the longitudinal coordinate (assumed to increase from the value zero at the stapes) and ω the radian frequency.[1] This function relates the pressure in the upper scala (scala vestibuli) to the transverse velocity of the basilar membrane at the *same* location along the BM. Write the relation between pressure $p(x, \omega)$ and BM velocity $v_{BM}(x, \omega)$ as[2]

$$p(x, \omega) = -1/2\, v_{BM}(x, \omega) Z_{BM}(x, \omega). \qquad (1)$$

In a classical model the pressure only depends on the *local* velocity. Furthermore, the BM impedance $Z_{BM}(x, \omega)$ is a *driving-point* impedance and obeys certain restrictions, the most important of these being that it corresponds to a causal impulse response. In fact, even its inverse has this property.

In a nonclassical model (examples: Steele *et al.*, 1993; Geisler and Sang, 1995; Wen and Boahen, 2003), the pressure at one location depends on the BM velocity at more than one location.[3] For instance, in a feed-forward model the pressure at location x depends on the velocity at the same location, as well as on the velocity at location $x - \Delta x$ where Δx is a positive constant—generally taken to be equal to several times the diameter of an outer hair cell. In that type of model the BM impedance plays an entirely different part. As a matter of fact, it can only be determined *after* the model equation has been solved, there is not an *a priori* governing impedance. In order to distinguish it, the BM impedance retrieved from the model response should then be called *effective* BM impedance. For a nonclassical model the following theorem has been proven (de Boer, 1997): a classical model that has the effective BM impedance of a nonclassical model as its impedance has the same response. At present it should be realized that this theorem is subject to a restriction. The effective BM impedance can, as said earlier, only be retrieved after the model equations have been solved. That solution will depend on the way the model is driven. In other words, the theorem is valid only when the same type of excitation of the two models (the classical and the nonclassical model) is considered.

II. ANALYSIS

In the present note we elaborate on this notion. We will ask and answer the question: does a nonclassical model behave differently for forward and reverse waves? Consider a feed forward model. Feed forward means that at a certain location an effect is felt from the velocity at a smaller value of x, i.e., at a location nearer to the stapes. Equation (1) will then read:

$$p(x, \omega) = -1/2\, v_{BM}(x, \omega) Z_{BM}(x, \omega)$$
$$- v_{BM}(x - \Delta x, \omega) B(x, \omega). \qquad (2)$$

In this equation the parameter $Z_{BM}(x, \omega)$ is the intrinsic impedance of the BM, $p(x, \omega)$ is the scala pressure at location x, $v_{BM}(x, \omega)$ the BM velocity at the same location and $v_{BM}(x - \Delta x, \omega)$ the BM velocity at a location lying Δx units more towards the stapes. When the interaction is like that of a stiffness, the parameter $B(x, \omega)$ (the symbol B is referring to "bridge", but it has the same dimension as Z_{BM} and thus the character of a transfer impedance) is positive (or negative) imaginary. The distance Δx is usually taken to be several tens of micrometers. We can only approach the concept of effective BM impedance when the model equation has been solved. Consider the case where the solution has appeared in the form of a wave with wave

[a]Electronic mail: e.d.boer@hccnet.nl

number k. That is, the BM velocity $v_{BM}(x,\omega)$ is assumed to vary with x as

$$\nu_{BM}(x,\omega) = \nu_0 \exp(-ikx). \tag{3}$$

For a forward wave k is positive. Substitute Eq. (3) in Eq. (2):

$$p(x,\omega) = -\nu_0 \exp(-ikx)(1/2Z_{BM}(x,\omega)$$
$$+ \exp(+ik\Delta x)B(x,\omega)). \tag{4}$$

Use the approximation $k\Delta x \ll 1$ (Δx is generally much smaller than the wavelength) and expand to the first order:

$$p(x,\omega) = -\nu_{BM}(x,\omega)(1/2Z_{BM}(x,\omega) + (1 + ik\Delta x)B(x,\omega)). \tag{5}$$

When the transfer impedance $B(x,\omega)$ is positive imaginary, Eq. (5) means that the (negative) imaginary component of $Z_{BM}(x,\omega)$ is made less negative by (the imaginary part of) $2B(x,\omega)$. At the same time a real component with the value $2ik\Delta xB(x,\omega)$ is added, see the last term in Eq. (5). When $B(x,\omega)$ is positive imaginary, this is a *negative resistance* which increases with increasing k. This analysis is true as long as $k\Delta x \ll 1$. Actually, when we consider Fig. 1 of de Boer *et al.* (2007), the magnitude of the wave number at the peak of the (measured) response is 13.8 (mm^{-1}). With Δx chosen as 0.070 (mm), the largest value used in Geisler and Sang (1995), the product $k\Delta x$ becomes equal to 0.97. Therefore, near the location of the response peak the first-order analysis ceases to be valid.

The modified pressure of Eq. (5) can be interpreted as the "generator" of the effective impedance of the nonclassical model. This is so because Eq. (5) describes the relation of pressure to velocity for a response of the form of $v_{BM}(x,\omega)$ as given by Eq. (3). A classical model with the quotient $-2p(x,\omega)/v_{BM}(x,\omega)$ from Eq. (5) as its BM impedance would yield exactly the same response. The influences on $Z_{BM}(x,\omega)$ as described above can thus be seen as determining the properties of the effective BM impedance $Z_{BM}^{eff}(x,\omega)$ with respect to the original impedance $Z_{BM}(x,\omega)$. In summary: with $B(x,\omega)$ positive imaginary, $Z_{BM}^{eff}(x,\omega)$ has a more positive imaginary part than $Z_{BM}(x,\omega)$—which causes the phase slope to increase—and $Z_{BM}^{eff}(x,\omega)$ has a negative-going real part than $Z_{BM}(x,\omega)$—which tends to produce amplification. Both are known properties of the response of a feed-forward model (Geisler and Sang, 1995).

When k is positive, Eq. (3) describes a forward wave. It is very simple to predict what will happen if a *reverse* wave is traveling in this model; we simply invert the sign of k. When $|k\Delta x| \ll 1$, Eq. (5) remains valid. The first effect is that the resistive component of $Z_{BM}(x,\omega)$ is increased by a positive quantity. This means that *attenuation* occurs instead of amplification. There is no (or a minimal) effect on the slope of the phase delay. Our analysis is approximate, nevertheless, the effects found will be the major effects in a feed-forward model for reverse *versus* forward waves. In summary: when a feed-forward model shows amplification for forward waves, it will show attenuation of reverse waves. In the feed-forward model the propagation velocity for both type of waves will be the same, the propagation velocity is de-

creased and the group delay increased with respect to the model without feedback.

In the case of *feed backward*, we only have to invert the sign of Δx. If we then assume $B(x,\omega)$ to be negative imaginary, we again have a model that can amplify forward waves. Under the same approximation, $|k\Delta x| \ll 1$, the phase slope is *diminished* for a forward wave—see Eq. (5) with negative imaginary $B(x,\omega)$ and negative Δx. In this case the effective stiffness associated with the transfer impedance $B(x,\omega)$ is increased. The rules for a reverse wave in this feed-backward model (in which a forward wave is amplified) are: the reverse wave is *attenuated* instead of amplified, and again the phase slope is minimally affected by inversion of the propagation direction.

In summary: when a *feed-forward* model is constructed so as to show amplification for forward waves, it will show attenuation of reverse waves. In the feed-forward model the propagation velocity is decreased and the phase slope increased with respect to the model without feedback, but the propagation velocities for forward and reverse waves will be the same. Furthermore, when in a *feed-backward* model the parameters are chosen such that the model shows amplification for forward waves, it will show attenuation of reverse waves. In the feed-backward model the propagation velocity is increased and the phase slope decreased with respect to the model without feedback, but the propagation velocities for forward and reverse waves will be the same.

III. COMBINATION OF FEED FORWARD AND FEED BACKWARD

The analysis above prompts us to consider the combination of feed forward, with a positive imaginary $B(x,\omega)$, and feed-backward, with a negative imaginary $B(x,\omega)$. Both types of feedback give amplification to a forward wave, and attenuation to a reverse wave. Their effects on the phase function tend to cancel, meaning that the phase stays close to that of the passive model, a desirable feature in view of the trends shown in experimental data on basilar-membrane movement in alive and dead animals. [4] In the model treated in the Titisee paper (de Boer and Nuttall, 2003) a combination of "normal" excitation, feed forward and feed backward was investigated. The set of coefficients giving the best fit between the measured and the computed response was obtained by a specialized solution procedure. From the present analysis it emerges that such an optimal solution would at least have one special feature, namely, that a component with positive Δx and a component with negative Δx, with opposite sign of the coefficient, should be involved. Seen as a function of x, the spatial convolution would include two terms with positive and negative Δx with opposite polarities. It remains to be seen whether a pattern with this characteristic will emerge with a more sophisticated deconvolution technique à la Titisee.

IV. CONCLUSION

In a feed-forward model designed to produce wave amplification for a forward wave, a reverse wave will be attenuated instead of amplified. The slope of the phase pattern is

2820   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Egbert de Boer: Non-classical models

(generally appreciably) enlarged for both types of waves. In a feed-backward model, again designed to produce wave amplification for a forward wave, a reverse wave is attenuated, and the slope of the phase pattern is smaller than in a model without feedback, for both types of waves.

ACKNOWLEDGMENT

[1] By the appearance of the radian frequency we limit ourselves to linear models here.

[2] The minus sign and the factor 1/2 are due to sign conventions. Upward movement of the BM is counted positive, $p(x,\omega)$ is the pressure in the upper channel and $-p(x,\omega)$ in the lower channel. Velocity and pressure should be interpreted as averages over the width of the BM.

[3] We tacitly consider models consisting of two compartments, the scala media being lumped into the cochlear partition in which the BM is the principal mechanical element. A general theory of three-compartment models has, to the knowledge of the author, not been developed yet.

[4] We have tested this prediction with a numerical solution and found it to be true. We did not describe and include this test to conserve space. The MATLAB program or a few representative figures can be obtained from the author.

de Boer, E. (**1997**). "Classical and non-classical models of the cochlea," J. Acoust. Soc. Am. **101**, 2148–2150.

de Boer, E., and Nuttall, A. L. (**2003**). "Properties of amplifying elements in the cochlea," in: *Biophysics of the Cochlea: From Molecules to Model*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 331–342.

de Boer, E., Nuttall, A. L., and Shera, C. A (**2007**). "Wave propagation patterns in a "classical" three-dimensional model of the cochlea," J. Acoust. Soc. Am. **121**, 352–362.

Geisler, C. D., and Sang, C. (**1995**). "A cochlear model using feed-forward outer-hair-cell forces.," Hear. Res. **86**, 132–146.

Steele, C. R., Baker, G., Tolomeo, J., and Zetes, D. (**1993**). "Electromechanical models of the outer hair cell," in: *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 207–214.

Wen, B., and Boahen, K. (**2003**). "A linear cochlear model with active bi-directional coupling," in: *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. **3**, Cancun, Mexico, pp. 2013–2016.

Contributions of individual components to the overall loudness of a multitone complex[a]

Lori J. Leibold,[b] Hongyang Tan, Samar Khaddam, and Walt Jesteadt
Boys Town National Research Hospital, Omaha, Nebraska 68131

The contributions of individual components to the overall loudness of a multitone complex were examined in a two-interval, loudness judgment task. Stimuli were five-tone complexes centered on 1000 Hz, with six different logarithmic frequency spacings, corresponding to bandwidths from 46 to 2119 Hz. Stimuli were presented for 300 ms (10 ms rise/fall). The overall level of the standard complex was fixed at 60 dB SPL (53 dB/component). Levels of the individual components of the comparison complex were selected at random from a rectangular distribution with a range of 20 dB. Perceptual weights were computed by calculating the point-biserial correlation between the difference in the level of each component across the two intervals and the subject's response. When all components fell within the same critical band, weights varied little across components. In contrast, the range of weights increased with increasing frequency separation, with more weight given to the lowest and highest frequency components. A relation between component weight and masked threshold was observed for all but the widest spacing condition. Perceptual weights and masked thresholds were largely in agreement with the loudness model [Moore *et al.*, J. Audio. Eng. Soc. **45**, 224–237 (1997)], except at the widest bandwidth. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2715456]

PACS number(s): 43.66.Cb, 43.66.Dc [AJO]   Pages: 2822–2831

I. INTRODUCTION

In their classic experiment, Zwicker *et al.* (1957) demonstrated that loudness does not generally depend on bandwidth when the bandwidth remains less than a single critical band and the total energy remains constant. In contrast, increasing the bandwidth beyond a single critical band results in a corresponding increase in loudness. This result is often summarized by saying that energy sums for components that are in the same critical band whereas loudness sums for components that are in separate critical bands. Subsequent studies have verified that the frequency separation of components plays an important role in the loudness of a complex sound (e.g., Scharf, 1961; Buus *et al.*, 1998).

If energy sums for components located within a critical band, the relative amplitudes of the individual components of a multitone complex should not affect a listener's loudness judgment, provided that all components are located in the same critical band and that the overall energy of the complex remains constant. However, when the bandwidth of the complex begins to exceed the critical band, we might expect that components will not contribute equally to the loudness of the complex. In support of this idea, Scharf (1962) reported that the spectral shape of a three-tone complex influenced loudness when the frequency spacing of the complex was greater than the critical band but not when the individual components fell within the same critical band.

Although loudness summation can be used to estimate the width of the critical band, any general model of loudness must take into account partial masking effects across several critical bands. Hübner and Ellermeier (1993), for example, observed that loudness summation does not hold under all conditions for two-tone complexes. The excitation-pattern model of loudness proposed by Moore *et al.* (1997) predicts masking between components, and a resulting reduction in loudness, even when the components are widely separated in frequency.

It is difficult to fully determine the effects of partial masking on loudness, because reduction in loudness is often used as a measure of partial masking (e.g., Fletcher and Munson, 1937; Scharf, 1964). Partial masking is commonly examined using loudness matching tasks in which subjects are asked to adjust the level of a target sound presented in isolation to the level of a target sound presented with a masker. When matching is repeated for various fixed levels of the target, the level of the target alone can be plotted as a function of the level of the target in the target-plus-masker interval at the point where the subject assigns equal loudness (reviewed by Gockel *et al.*, 2003). It is important to note, however, that this task requires subjects to consider only the loudness of the target and it is not clear whether subjects can successfully attend to the target and ignore the masker when they are asked to do so.

A complete understanding of loudness perception cannot be achieved without knowing how individual components of a complex sound contribute to its overall loudness. The contribution of individual components to the discrimination of multitone complexes has been assessed by the measurement of perceptual weights (e.g., Doherty and Lutfi, 1996; Korte-

---

[a]Portions of these results were presented at the Acoustical Society of America meeting in Vancouver, BC in May 2005 [Leibold et al., J. Acoust. Soc. Am. 117, 2597 (2005)].
[b]Current affiliation: Department of Allied Health Sciences, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599; electronic mail: leibold@med.unc.edu

2822   J. Acoust. Soc. Am. **121** (5), May 2007   0001-4966/2007/121(5)/2822/10/$23.00   © 2007 Acoustical Society of America

kaas *et al.*, 2003). Weights can be computed from the correlation between the subject's responses and random variations in the magnitude of each feature in a complex stimulus on individual trials (Berg, 1989; Lutfi, 1989; 1995; Richards and Zhu, 1994). Higher relative weights indicate features that have a greater influence on the response. This measurement technique is now well established in the literature on intensity discrimination and masking, but has only recently been applied to loudness (Leibold and Werner, 2003). In addition to providing information about the relative contributions of individual components of a complex sound during overall loudness judgments, comparing the perceptual weights individuals apply in a condition where loudness is believed to be additive and in a condition where it is not additive offers an opportunity to examine these conditions more closely and to develop procedures that provide more accurate descriptions of loudness summation.

The purpose of this study was to determine the contribution of individual components to judgments of overall loudness in the context of a band-widening experiment (e.g., Zwicker *et al.*, 1957). To accomplish this goal, the perceptual weight subjects assign to the different spectral components of a five-tone complex was measured during a loudness judgment procedure. In addition, the relation between the perceptual weight and audibility of individual components was examined. The data were then used to test the model of loudness proposed by Moore *et al.* (1997).

II. EXPERIMENT 1: LOUDNESS MATCHES BETWEEN A MULTITONE COMPLEX AND A 1000-HZ TONE

The goal of the first experiment was to verify the relation between the bandwidth of a multitone complex and its perceived loudness for stimuli and subjects to be used in the perceptual weights experiment. The level of a 1000-Hz tone equal in loudness to a five-tone complex was estimated for each of six frequency-spacing conditions. Following Zwicker *et al.* (1957), the prediction was that loudness would be independent of frequency spacing when all components fell within one critical band. As frequency spacing increased beyond the critical band, however, the prediction was that subjects' loudness estimates would increase.

A. Subjects

The subjects were seven adults (18–40 years), including five paid volunteers and authors SK (S1) and LL (S3). All subjects had normal hearing sensitivity, with thresholds for 200-ms tones in quiet of 20 dB SPL or better at 500, 1000, 2000, and 4000 Hz in both ears, when measured using a two-interval, forced-choice (2IFC) adaptive procedure.

B. Stimuli and equipment

The stimuli were five-tone complexes centered on 1000 Hz and a 1000-Hz pure tone. All stimuli were presented for 300 ms, including 10-ms, $\cos^2$, onset/offset ramps. Six different logarithmic spacing conditions were used, ranging from a frequency ratio of 1.012 between adjacent components to a frequency ratio of 1.586. This spacing corresponds to bandwidths from 46 to 2119 Hz. For the two

narrowest spacing conditions, all five components fall within a single critical band at 1000 Hz, based on the equivalent rectangular bandwidth (ERB) measure proposed by Glasberg and Moore (1990). For the two middle spacing conditions, the five components cover ranges of 1.7 and 3.5 ERBs. For the two widest spacing conditions, the ranges were 6.9 and 13.7 ERBs and each component fell in a different critical band. The overall level of each complex was 60 dB SPL (53 dB/component).

Stimuli were digitally generated at a 50-kHz sampling rate using an array processor (TDT AP2) and were played through 16-bit digital-to-analog converters (TDT DD1). The stimuli were then low-pass filtered at 20 kHz (TDT FT6) and attenuated (TDT PA4). The output of the attenuator was fed to a headphone buffer (TDT HB6), then to a remote passive attenuator in the sound-treated room, and finally to the left ear of the subject through a Sennheiser HD 250 Linear II headphone.

C. Procedure

Subjects were tested as a group of three (S1, S6, S7), and two groups of two (S2, S5 and S3, S4). Each subject was tested in a separate section of a double-walled Industrial Acoustics sound-treated room that had single walls between sections. The subject groups were run in 2-h sessions, including regular breaks.

Loudness matches were made between the multitone complex for each frequency spacing condition and a 1000-Hz tone using a two-track, 2IFC adaptive loudness matching procedure (Jesteadt, 1980). The 1000-Hz tone and the complex were presented in random order on each trial. Subjects were instructed to select the interval that contained the louder sound and the level of the 1000-Hz tone was adapted based on those responses. The adaptive procedure used decision rules that estimated the 71% point of the psychometric function on one track and 29% on the other track (Levitt, 1971), using starting levels of 70 and 50 dB SPL, respectively. For each track, the initial step size was 4 dB, followed by a step size of 1 dB after the fourth reversal. The 71% or 29% point was estimated by computing the average of all reversal points for the 1-dB step size.

Subjects completed two blocks of 100 trials (50 trials/track) for each frequency spacing condition. Test order was randomized across conditions for each block of trials. Each trial consisted of a 300-ms warning interval and two 350-ms observation intervals separated by 300 ms. There was a 300-ms interval before the beginning of the next trial. Visual markers for the warning and observation intervals appeared on the message window of a small keypad that subjects used to indicate their responses. No response feedback was provided in any of the spacing conditions.

D. Results

For each block of trials, the loudness match for each subject was obtained by averaging across both adaptive tracks. The data reported here are the estimates averaged across both blocks of trials. The group mean level of the 1000-Hz tone judged equal in loudness to the multitone com-

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Leibold *et al.*: Loudness weights   2823
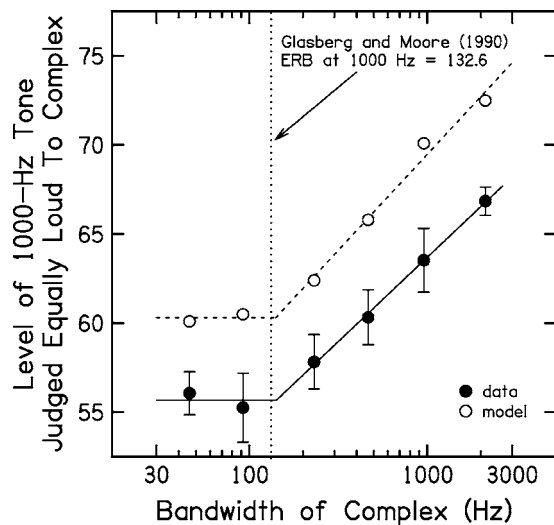
FIG. 1. Mean level, averaged across seven subjects, of a 1000-Hz tone judged equal in loudness to the multitone complex at 60 dB SPL is shown (closed circles) as a function of the bandwidth of the complex. Error bars are ±1 S.E. of the mean level across subjects. The dotted vertical line indicates the equivalent rectangular bandwidth (ERB) at 1000 Hz (Glasberg and Moore, 1990). Estimates of the mean level of a 1000-Hz tone judged equal in loudness to the multitone complex at 60 dB SPL provided by the Moore *et al.* (1997) model are provided (open circles).

plex at 60 dB is shown by the circles in Fig. 1, plotted as a function of the bandwidth of the complex. Error bars represent ±1 S.E. of the mean level across subjects. The solid lines represent the best two-line fit to the data, with the restriction that one of the two lines was horizontal. The horizontal line represents the hypothesis that loudness will be constant when all components are within a single critical band. It was fitted to the loudness matches obtained for complexes with bandwidths of 46 and 92 Hz. The sloping line was fitted to the loudness matches obtained for complexes with bandwidths of 231, 456, 956, and 2119 Hz. The vertical dotted line indicates the ERB at 1000 Hz (Glasberg and Moore, 1990). Estimates of the mean level of a 1000-Hz tone judged equal in loudness to the multitone complex at 60 dB SPL provided by the Moore *et al.* (1997) loudness model are shown by the open circles. The dashed lines represent the best two-line fit to the model predictions, as described above for the best two-line fit to the data.

E. Discussion

The data are consistent with the assumption that the loudness of the multitone complex is independent of bandwidth when components are located within a critical band. The two complexes with the closest loudness matches were those with 46- and 92-Hz bandwidths. The average loudness match for these complexes was 55.7 dB SPL. As expected, the loudness of the complex increased as bandwidth increased beyond the critical band. The intersection point of the two lines was 141.5 Hz, a value that is similar to the ERB estimate of 132.6 at 1000 Hz (Glasberg and Moore, 1990).

Overall, these results are consistent with previous reports (e.g., Zwicker *et al.*, 1957) and support the idea that components that are in the same critical band are summed

before conversion to loudness whereas components that are in separate critical bands are converted to loudness, and then summed. Some interaction among components occurs in both cases. If complete energy summation occurred within a critical band, the two narrowest bandwidths should be equal in loudness to a 60-dB tone. The average loudness match from our subjects in these two conditions, however, indicates that the loudness of the complex was approximately 4 dB less than a 60-dB tone when all components were expected to be located in one critical band. Other studies have found less than complete energy summation in similar conditions (e.g., Scharf, 1961), although the difference observed here is larger than previously reported. A significant portion of the 4-dB difference between the level of the tone and the level of the narrow complexes can be attributed to the performance of two listeners. The use of two adaptive tracks provides estimates of the 29% and 71% points on the psychometric function (PF) for the loudness-matching task and thus an estimate of the slope of the PF. For 5/7 subjects (S1, S3, S4, S5, and S6), the mean difference between the 29% and 71% points ranged from 1.7 to 9.1 dB across the six spacing conditions. For the remaining two subjects (S2 and S7), the mean difference between the 29% and 71% points ranged from 17.3 to 23.3 dB across the six spacing conditions. There was a tendency for the shallower PFs to have lower midpoints. The basis for the erratic performance shown by these two listeners might reflect lack of training and relatively small amount of data, 200 trials per condition, contributing to each estimate. The data for these two subjects were not excluded from the study because their data were consistent with the group data for the later experiments.

Interaction among components also occurs when the components are spread across critical bands. For the two conditions with frequency ranges of 956 and 2119 Hz, for example, the spacing between adjacent components increases from 1.4 ERB, on average, to 2.7 ERB. The simplest interpretation of loudness summation as proposed by Zwicker *et al.* (1957) would suggest that loudness would be at its maximum for the 956-Hz range, where all components were in separate critical bands, and that no further increase in loudness would be observed when the range increased to 2119 Hz. For the four widest complexes, the data shown in Fig. 1 indicate that loudness increased at a rate of approximately 2.8 dB/octave with increasing bandwidth. The model proposed by Moore *et al.* (1997) predicts an increase of 3.2 dB per octave for these conditions, with no decrease in slope at the larger bandwidths. The model assumes partial masking across several critical bands before conversion to loudness and predicts differences in loudness summation for stimuli within and between critical bands based solely on differences in masking as a function of frequency spacing.

III. EXPERIMENT 2: PERCEPTUAL WEIGHTS FOR INDIVIDUAL COMPONENTS OF A MULTITONE COMPLEX

The goal of the second experiment was to estimate perceptual weights for individual frequency components of the multitone complex as a function of the frequency spacing between components. If the level of a particular component

is correlated with the loudness judgment, it indicates that information in that frequency region contributes to the overall loudness of the complex. The approach was to measure perceptual weighting functions using a two-interval, loudness judgment task, similar to the task used by Doherty and Lutfi (1996) for measuring perceptual weights during intensity discrimination.

A. Subjects, stimuli, equipment, and procedure

The subjects, stimuli, equipment, and procedure were identical to those described in Experiment 1, with the following exceptions. Five-tone complexes were presented in both observation intervals. The overall level of the standard complex was fixed at 60 dB SPL and all five tones were equal in level. For the comparison complex, levels of the individual components were selected at random from a rectangular distribution of levels with a range of 20 dB and a step size of 1 dB, centered on 53 dB. The standard and comparison complex were presented in random order on each trial. The subject's task was to indicate which interval contained the louder complex.

Ten 100-trial blocks were completed for each frequency-spacing condition. As in Experiment 1, trials consisted of a 300-ms warning interval and two 350-ms observation intervals separated by 300 ms. Visual markers indicated the warning and both observation intervals. Subjects were not provided with response feedback.

B. Results

Following Lutfi (1995), weights for individual subjects were determined from the following linear regression model,

$$D = \Sigma w_i x_i + C,$$

where D is the subject's response, x_i is the difference in level of the ith component across the two intervals, w_i is the weight applied to the ith component, and C is a constant. The constant C corresponds to the accumulation of additive internal noise occurring both before and after weights are applied.

This is essentially the procedure used in many previous studies that have examined perceptual weights during auditory tasks (e.g., Lutfi, 1995; Richards and Zhu, 1994; Stecker and Hafter, 2002; Stellmack et al., 1997; Willihnganz et al., 1997), except that subjects were asked to respond based on loudness. The regression coefficients can be viewed as unnormalized or "beta" weights.

A significant relationship between component level and response was obtained ($p < 0.05$) for each component at all bandwidths for subjects S3, S4, S6, and S7. That is, all components contributed to the loudness of the multitone complex in these subjects. For S1, no significant relationship was observed for the second component in the 92-, 231-, and 2119-Hz conditions. For S2, no significant relationship was observed for the second component in the 46-, 92-, and 956-Hz conditions as well as for the first component in the 456-Hz condition. For S5, no significant relationship was observed for the first component in the 2119-Hz condition.
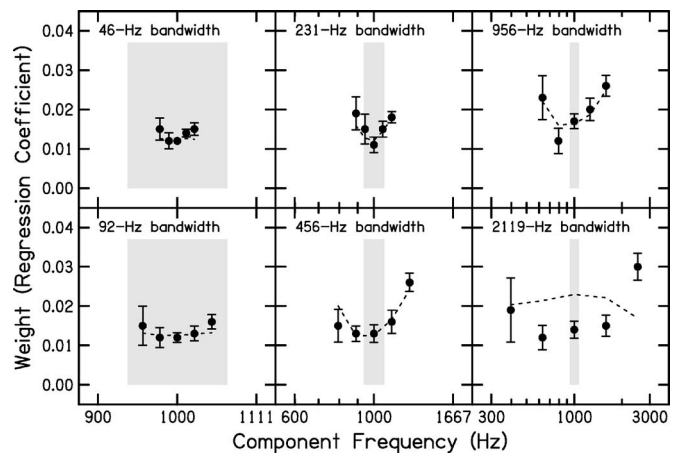
FIG. 2. Mean weights for the individual components of the multitone complex plotted as a function of the bandwidth of the complex (closed circles). Error bars are ±1 S.E. of the mean weight across subjects. The shaded area represents the ERB centered on 1000 Hz. Model predictions for each spacing condition are shown by the dashed lines.

Figure 2 shows mean weight as a function of component position for each bandwidth. Data for each spacing condition (closed circles) are provided in separate panels. The shaded area represents the ERB centered at 1000 Hz. Error bars represent ±1 S.E. of the mean weight across subjects. When all components fell within the same critical band, weights varied little across components. In contrast, the range of weights increased with increasing frequency separation, with increasing weight being given to the lowest and highest frequency components.

A repeated-measures analysis-of-variance (ANOVA) confirmed the trends observed in Fig. 2. All of the effects in the bandwidth × component analysis were significant: Bandwidth [$F(5,30) = 6.1$; $p < 0.01$], component [$F(4,24) = 4.0$; $p < 0.05$], and bandwidth × component $F(20,120) = 2.0$; $p < 0.05$]. The bandwidth × component interaction suggests that differences in weight across components are not equivalent across the six spacing conditions. This interaction was examined further by performing a one-way ANOVA for each spacing condition. This analysis revealed a significant main effect of component for the three widest conditions: 456-Hz bandwidth [$F(4,24) = 6.4$; $p < 0.01$], 956-Hz bandwidth [$F(4,24) = 4.3$; $p < 0.01$], and 2119-Hz bandwidth [$F(4,24) = 2.9$; $p < 0.05$]. In contrast, no significant main effect of component was found for the three narrowest conditions: 46-Hz bandwidth [$F(4,24) = 0.8$; $p = 0.6$], 92-Hz bandwidth [$F(4,24) = 0.6$; $p = 0.7$], and 231-Hz bandwidth [$F(4,24) = 0.8$; $p = 0.1$].

Individual differences in weight are evident by the error bars shown in Fig. 2. These differences were largest for conditions with the greatest spacing between components. For example, large differences across subjects are apparent for weights computed for the complex with the 2119-Hz bandwidth, especially for the lowest and highest frequency components. In contrast, weights were more similar across subjects for the complex with the 46-Hz bandwidth.

Weights predicted by the Moore et al. (1997) loudness model are also shown in Fig. 2 (dashed lines). Weights were determined for the loudness model by computing the overall

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Leibold et al.: Loudness weights   2825
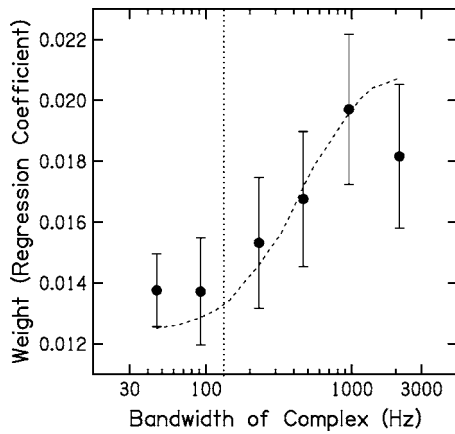
FIG. 3. Mean weight across the five components of the multitone complex plotted as a function of the bandwidth of the complex (closed circles). Error bars are ±1 S.E. of the mean weight across subjects. The vertical dotted line indicates the ERB centered on 1000 Hz. Model predictions for each spacing condition are shown by the dashed lines.

loudness of the two multitone complexes actually presented on each trial and having the model vote for the louder of the two. The decisions were then correlated with the level of each of the five tones for the variable-level stimulus. Data files from Experiment 2 contained lists of 1000 trials for each of the six bandwidths with at least two lists per bandwidth. Each trial specified the levels of each of the five tones. Weights for the model were estimated by computing a set of five weights for each bandwidth using each of the twelve independent lists of tone levels. The twelve values of the weight of each of the five tones were then averaged to obtain an estimate of the weight based on 12 000 trials. The 30 weights obtained for the model exceeded the observed weights by 35% on average, because the model always voted consistently. If each set of weights was normalized, the difference between observed and normalized weights would have been corrected as part of the normalization process. All model weights shown in Fig. 2 were divided by 1.35 to bring them in line with the data.

Figure 3 provides the average weight across all subjects and components for each spacing condition, plotted as a function of the bandwidth of the complex. Error bars represent ±1 S.E. of the mean weight across subjects. Consistent with the loudness matches obtained in Experiment 1, average weight varied little across the two conditions in which all components were located in a single critical band, but generally increased with increasing frequency separation.

To obtain a smooth function showing the relation predicted by Moore *et al.* (1997) between bandwidth and the average weight across the five tones in Fig. 2, the original set of weights at each of six bandwidths was expanded by computing fifteen additional sets of logarithmically spaced tones at intermediate bandwidths. For each of these new bandwidth conditions, twelve sets of five weights were obtained for the model using the twelve independent trial lists described earlier and the sixty weights were then averaged and corrected by dividing by 1.35. The model weight estimates generated in this way are shown by the dashed line in Fig. 3.

C. Discussion

Table I presents the percentage of trials where the complex with levels that varied randomly was selected as the louder of the two sounds. Subjects found the complex with levels that varied to be louder than the complex with fixed levels, but the difference decreased with the increased spacing of the components. This result provides an additional source of information regarding the effective transition from energy summation to loudness summation with increasing bandwidth. Subjects found the interval with five tones that varied randomly in level to be louder on average than the interval with five tones fixed at an overall level of 60 dB SPL because the variable levels were drawn from a rectangular distribution expressed in decibels. Given the parameters that were used, the average power of the five-tone complex with variable levels was equivalent to 63.6 dB SPL, whereas the five-tone complex with fixed levels was presented at 60 dB SPL. For conditions where the components were closely spaced in frequency, the traditional view is that the total power is converted to loudness and the complex with variable levels should be louder on average. If individual components are converted to loudness first and then summed, however, the effect of variable levels should be reduced because the conversion to loudness is compressive. If we assume a power law with an exponent of 0.3, we predict that the five-tone complex with components widely spaced in frequency will be equivalent in loudness, on average, to a tone presented at 61.2 dB SPL. Thus, the reduction in the proportion of cases where the variable-level complex was found to be louder as a function of component spacing is consistent with greater loudness summation at wider bandwidths. Table I also shows the percentage of trials where the Moore *et al.* (1997) loudness model would predict that the complex with levels that varied randomly was the louder of

TABLE I. Percentage of trials in Experiment 2 where subjects selected the interval with the jittered levels. The column on the right shows the percentage of trials the Moore *et al.* (1997) loudness model predicts the interval with the jittered levels is louder.

Condition	Subject							Mean	Model
	S1	S2	S3	S4	S5	S6	S7		
46-Hz bandwidth	72.8	63.2	71.0	82.4	79.1	78.7	72.7	74.3	83.2
92-Hz bandwidth	70.1	64.7	65.3	83.8	80.8	76.9	71.9	73.4	83.6
231-Hz bandwidth	71.1	66.6	60.1	79.3	76.4	78.0	63.4	70.7	80.7
456-Hz bandwidth	58.7	57.2	65.9	86.6	70.8	68.1	61.9	66.9	74.4
956-Hz bandwidth	55.3	54.7	53.0	67.4	61.2	55.3	50.9	56.8	68.9
2119-Hz bandwidth	57.5	57.3	60.7	60.1	63.4	51.4	51.0	57.3	64.5

Leibold *et al.*: Loudness weights

the two sounds. Loudness judgments by the model show a stronger preference for the variable-level complex than the judgments by the subjects in all conditions, with a constant preference in the narrowest bandwidth conditions that declined with increased frequency spacing. The model finds the variable-level complex to be the louder of the two more often than the subjects do because it makes consistent decisions from trial to trial and is more sensitive to the differences between the fixed- and variable-level complexes.

The audibility of the individual components might have contributed to the pattern of weights shown in Fig. 2. A second, related factor is the extent to which multiple components fall into the same critical band. Both factors are incorporated in the model of loudness perception proposed by Moore *et al.* (1997). As shown by the dashed lines in Figs. 2 and 3, the model weights are consistent with complete energy summation for the narrowest bandwidths (46 and 92 Hz). The components contribute equally to the overall loudness of the complex and the effect of varying the level of any one component is the same for all components. Note, however, that the loudness model makes no distinction between loudness summation within a critical band and across critical bands. Instead, the Moore *et al.* (1997) loudness model computes excitation patterns and converts them to specific loudness. A different pattern is observed in the model predictions for the middle bandwidths (231, 456, and 956 Hz). For these conditions, the loudness perception model predicts differences in the contributions of individual components. Specifically, the model predicts greater weight for components on the edges of the complex (1 and 5) relative to the three middle components (2, 3, and 4). Finally, the model weight predictions are markedly different for the widest bandwidth (2119 Hz), where the influence of varying the middle components is somewhat greater than the influence of varying the edge components.

It is not clear why the model assigned greater weight to the middle components than to the edges in this condition. Increasing the bandwidth from 956 to 2119 Hz in small steps results in an orderly shift from greater weight on the edges to a relatively flat pattern to the pattern shown for the 2119-Hz condition. Shifting the components up in frequency while holding bandwidth constant does not have the same effect. It is the bandwidth rather than the specific frequencies that produces this result. The observed weights differ markedly from the weights predicted by the model only in this condition.

The pattern of weights observed for the most widely spaced condition are in agreement with those reported by Kortekaas *et al.* (2003) in a level discrimination task using seven-component multitonal stimuli. At levels comparable to those used here and with each component located in a separate critical band, Kortekaas *et al.* observed that listeners assigned the greatest weight to the highest frequency component. This weighting pattern persisted even when flanking bands of noise were introduced to mask the spread of excitation. Kortekaas *et al.* suggested that listeners might have chosen the interval that had the highest partial loudness for any single frequency component.

The existence of substantial individual differences across subjects and the deviation from the model predictions

in the most widely spaced condition suggests that central auditory processes contribute to how subjects assign loudness judgments when components are widely spaced. The model of loudness perception proposed by Moore *et al.* (1997) incorporates assumptions about the audibility of individual components, but makes the same assumptions for all subjects, of course. To explore the relation between perceptual weight and audibility in individual subjects, masked detection thresholds were obtained for each component in the presence of the remaining components.

IV. EXPERIMENT 3: AUDIBILITY OF INDIVIDUAL COMPONENTS

The goal of the third experiment was to measure the audibility of individual components within the multitone complex and to determine the relation between the perceptual weight and audibility of individual components as a function of frequency separation for individual subjects.

A. Subjects, stimuli, equipment, and procedure

Data were collected from the subjects described in Experiments 1 and 2, using the same stimuli and equipment. Masked detection thresholds were obtained for each of the five tones used in the perceptual-weights task, using the remaining four tones as the masker. The overall level of the four masker components was 60 dB SPL (54 dB/component).

Thresholds were estimated using a 2IFC, adaptive procedure that estimated the 71% point on the psychometric function (Levitt, 1971). The starting level of the adaptive track was 70 dB. The initial step size was 4 dB, followed by a step size of 2 dB after the fourth reversal. Each trial consisted of a 300-ms warning interval and two 350-ms observation intervals. The interstimulus interval was 500 ms. A 300-ms feedback interval followed the two observation intervals. During the feedback interval, a message window indicated the interval that contained the signal. Two 100-trial blocks were run for each component in each frequency-spacing condition. Thresholds for each block were computed by calculating the average of all reversal points for the 2-dB step size.

B. Results

Mean masked thresholds across subjects as a function of component frequency, with data for each spacing condition (closed circles) in separate panels, are shown in Fig. 4. As in Fig. 2, the shaded area represents the ERB centered at 1000 Hz and error bars represent ±1 S.E. of the mean threshold across subjects. Masked thresholds predicted by the Moore *et al.* (1997) loudness model are shown by the dashed lines. These threshold estimates were obtained by finding the level at which the partial loudness of the signal component was equal to 2 phons. Predicted thresholds underestimate observed thresholds, but the model provides a good account of the effects of component position and component spacing in most conditions. The form of the predicted thresholds deviates from the data only in the widest bandwidth condition, where individual differences are also large.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Leibold *et al.*: Loudness weights   2827
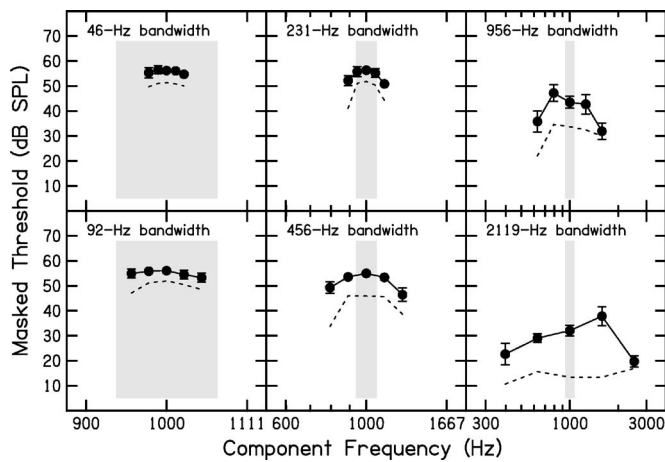
FIG. 4. Mean masked thresholds for the individual components of the multitone complex plotted as a function of the bandwidth of the complex (closed circles). Error bars are ±1 S.E. of the mean threshold across subjects. The shaded area represents the ERB centered on 1000 Hz. Model predictions for each spacing condition are shown by the dashed lines.

All of the data are represented as masked thresholds in dB SPL because it is the most consistent approach across the six bandwidth conditions. However, for conditions in which all components were located in the same critical band, the detection of one component in the presence of the others can be viewed as an intensity discrimination task in which the four fixed-level components represent a 60 dB SPL pedestal ($10 \log I$) and the signal component represents an intensity increment ($10 \log \Delta I$). In that framework, the mean thresholds for the 46- and 92-Hz bandwidth conditions correspond to values of ΔL (i.e., $10 \log [(I+\Delta I)/I]$) of 1.4 and 1.2 dB, respectively. It is not surprising that thresholds varied little across individual components. The range of thresholds increased, however, as the frequency separation between components increased. For example, mean thresholds varied across components over a range of less than 2 dB (54.7–56.5 dB SPL) for the 46-Hz bandwidth condition compared to a range of approximately 18 dB (19.7–37.8 dB SPL) for the 2119-Hz condition. In particular, the lowest and highest components became more audible with increasing frequency separation.

A repeated-measures ANOVA indicated significant effects of bandwidth [$F(5,30)=143.4$; $p<0.01$], component [$F(4,24)=11.9$; $p<0.01$], and bandwidth $\times$ component [$F(20,120)=14.2$; $p<0.01$]. The interaction reflects significant differences in threshold across components for the four widest spacing conditions: 231-Hz bandwidth [$F(4,24)=6.5$; $p<0.01$], 456-Hz bandwidth [$F(4,24)=6.8$; $p<0.01$], 956-Hz bandwidth [$F(4,24)=9.8$; $p<0.01$], and 2119-Hz bandwidth [$F(4,24)=6.7$; $p<0.01$]. No significant differences across component thresholds were found for either the 46-Hz bandwidth [$F(4,24)=0.9$; $p=0.5$] or the 92-Hz bandwidth condition [$F(4,24)=2.0$; $p=0.1$]. These results mirror those obtained for the weights, with the exception of a significant difference in threshold, but not weight, for the 231-Hz bandwidth condition.

Significant masking of each component by the other components is observed even in conditions where the components are widely spaced. Figure 5 shows the average
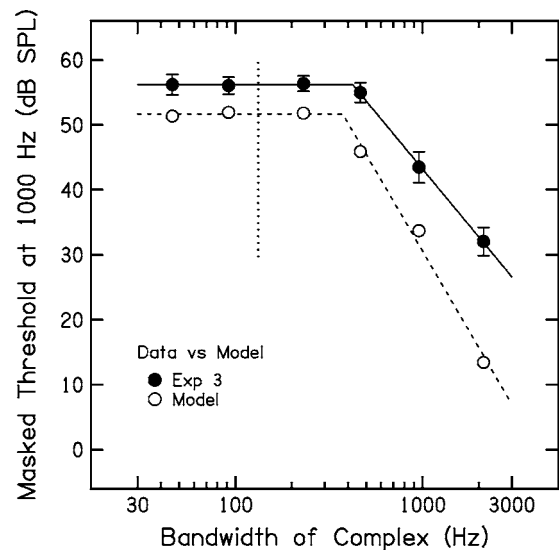
FIG. 5. Average masked threshold of the 1000-Hz component (closed circles) plotted as a function of the bandwidth of the complex. Error bars are ±1 S.E. of the mean threshold of the 1000-Hz tone across subjects. The dotted vertical line indicates the ERB at 1000 Hz. Model predictions for each spacing condition are provided (open circles).

masked threshold for the 1000-Hz component, which is the only frequency that occurs in all bandwidth conditions, as a function of bandwidth. Masked thresholds predicted by the Moore *et al.* (1997) loudness model are shown by the dashed lines. Although the model assumes masking across multiple critical bands, the model underestimates masked thresholds to a greater degree as the separation between tonal components increases.

Table II shows correlations of the average masked threshold of each component in the presence of the others, for individual subjects, with the loudness matching results for those subjects in Experiment 1 as well as with the average perceptual weights for those subjects in Experiment 2. For the three narrowest bandwidths, subjects with less masking of one component by another tended to match the five-tone complex to a more intense tone in Experiment 1 and their decisions were more highly correlated with the levels of the individual tones in Experiment 2. A similar trend was

TABLE II. Correlation across subjects of the mean masked threshold for each component, as measured in Experiment 3 with loudness data from Experiment 1 and weights from Experiment 2. Because there were only seven subjects, a correlation coefficient of 0.754 is required for significance. Significant correlations are marked by asterisks.

Condition	Correlation with loudness match in Experiment 1	Correlation with average weight in Experiment 2
46-Hz bandwidth	−0.759*	−0.815*
92-Hz bandwidth	−0.876*	−0.861*
231-Hz bandwidth	−0.945*	−0.843*
456-Hz bandwidth	−0.549	−0.479
956-Hz bandwidth	−0.480	−0.833*
2119-Hz bandwidth	0.040	−0.199
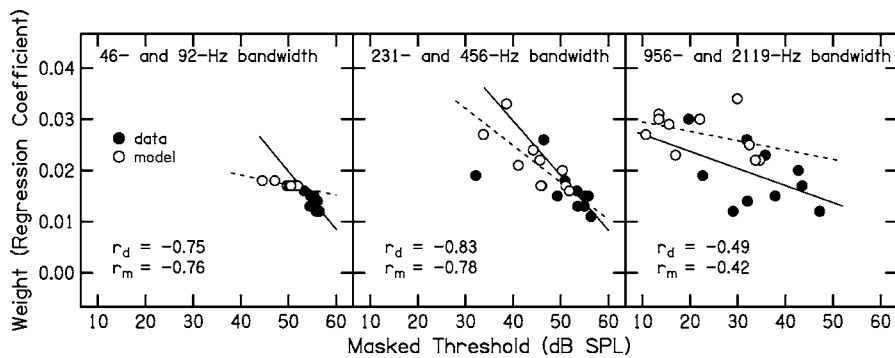
FIG. 6. Scatterplots of mean weight as a function of mean masked threshold for each component at the two narrowest (46- and 92-Hz bandwidth), two middle (231- and 456-Hz bandwidth) and two widest (956- and 2119-Hz bandwidth) conditions. The correlation coefficients r_d and r_m are provided within each panel for each linear regression, representing the data and model estimates, respectively.

observed at all but the widest bandwidth, but was not statistically significant in the 456- and 956-Hz bandwidth conditions.

Figure 6 shows the relation between perceptual weight and masked threshold for each component at the two narrowest bandwidths (46 and 92 Hz), the two middle bandwidths (231 and 456 Hz), and the two widest bandwidths (956 and 2119 Hz). Perceptual weights and masked thresholds are shown for the mean across subjects (closed circles) and for the model (open circles). The lines shown in each panel represent the best linear least squares fits when all 10 data points (2 bandwidths ×5 components) are included in a single analysis. The correlation coefficients within each panel, r_d and r_m, describe the relation between threshold and perceptual weight for the data and the model, respectively. The results shown in Fig. 6, where threshold and weight data are averaged across subjects before computing a correlation coefficient, are comparable to those shown in Table II, where thresholds and weights are averaged across tone position before computing a correlation coefficient. The relation between threshold and perceptual weight decreases as the bandwidth increases.

C. Discussion

Results from Experiment 3 are consistent with results from the two earlier experiments and support the hypothesis that the contribution of individual tonal components to the overall loudness of a five-tone complex is related to the audibility of the individual components at all but the widest bandwidth. The Moore *et al.* (1997) model incorporates audibility into its calculation of specific loudness, which is then integrated to arrive at overall loudness. The model provides a good account of perceptual weights at all but the widest bandwidth in Experiment 2 and of thresholds at the narrower bandwidths in Experiment 3. In their review of use of the model to predict masked thresholds, Moore *et al.* assumed a detection criterion of 2 phons in most cases, but increased the detection criterion to 8 phons to provide a better account of certain data. A more optimum fit of the model to the data in Figs. 4 and 5 could be achieved by assuming a criterion of 13 phons, but it seems unlikely that subjects had an unusually high detection criterion for these stimuli. It is more likely that the model underestimates the spread of excitation or the contribution of excitation to masking. Alternatively, the increasing discrepancy between actual thresholds and

model predictions with increasing bandwidth might reflect an increased contribution of central auditory processes with increasing bandwidth.

Although the model underestimates masked thresholds in the wide bandwidth conditions, the data shown in Fig. 5 demonstrate that the model and the data are in agreement with respect to the breakpoint in the function. The breakpoint represents an estimate of the critical band using masking rather than loudness as a measure of performance. Although the breakpoints for the model and the data for the loudness matches obtained in Experiment 1 (Fig. 1) are in good agreement with the ERB measure of the width of the critical band (Glasberg and Moore, 1990), the breakpoints in Fig. 5 are significantly larger than the ERB. Most studies that have measured the critical band using masking have varied the bandwidth of a noise masker while holding spectrum level constant (e.g., Bernstein and Raab, 1990; Hall *et al.*, 1984; Schaeffer *et al.*, 1950; Schooneveldt and Moore, 1989; Spiegel, 1981) rather than holding total masker power constant as was done here. Those studies report a wide range of bandwidths, some narrower than the ERB and some wider. When the Moore *et al.* (1997) model is used to estimate masked thresholds for a tone in constant-spectrum-level noise of varying bandwidth, the breakpoint is similar to that shown in Fig. 5.

Figure 6 provides a summary of the relation between audibility and perceptual weight. For the narrowest bandwidth conditions (46 and 92 Hz), a strong correlation was found for both the data ($r=-0.75$) and the model ($r=-0.76$). With 8 degrees of freedom, a correlation with an absolute value $\geqslant 0.765$ is significantly different from zero at the $p < 0.01$ level (Kleinbaum *et al.*, 1998), suggesting some degree of independent processing of tonal components. For the widest bandwidth conditions (956 and 2119 Hz), the perceptual weights varied across subjects and were not related to threshold.

The lack of a relation between component audibility and contribution to overall loudness in the widest bandwidth condition is illustrated in Fig. 7, which presents data for individual tonal components and individual subjects in that condition. The lack of significant correlation is not surprising, given that all of the components are well above threshold. Predictions of the model for that condition, shown by the open symbols in Fig. 7, show a narrow range of both thresholds and weights.

Different subjects use different rules for loudness sum-

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Leibold *et al.*: Loudness weights   2829
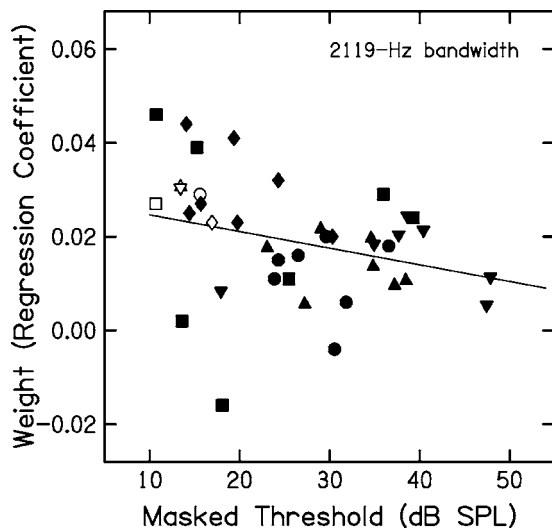
FIG. 7. Scatterplot of weight as a function of masked threshold for all subjects and components for the 2119-Hz bandwidth condition. Data for each component are indicated by separate symbols (■=397 Hz, ● =630 Hz, ▲=1000 Hz, ▼=1586 Hz, ◆=2516 Hz). The solid line represents the best linear squares fit to all 35 data points. Open symbols show weight as a function of masked threshold for each component as predicted by the model.

FIG. 8. Mean coefficient of contingency plotted as a function of the bandwidth of the complex showing the accuracy of prediction of subjects' decisions in Experiment 2 based on weights obtained from the data (closed circles) or on loudness estimates obtained from the model (open circles). Error bars are ±1 S.E. of the mean coefficient across subjects.

mation in the widest bandwidth condition and measurement of perceptual weights provides a means of assessing those individual differences. To obtain a measure of the information contained in the perceptual weights, the weights for each subject for each bandwidth condition in Experiment 2 were used to predict subjects' trial-by-trial decisions. The level of each component tone was multiplied by the perceptual weight associated with that component and the sum of those five values was compared to the corresponding sum for the matching stimulus that consisted of the same five tones that were all equal in level. If the sum for the variable-level stimulus was higher, the subject was predicted to vote that the variable-level stimulus was louder. The estimates of loudness obtained from the loudness model were also used to predict each subject's decisions. In this case, if the loudness associated with the variable-level stimulus was estimated to be greater than the loudness associated with the matching stimulus, the subject was predicted to vote that the variable-level stimulus was louder. The accuracy of both sets of predictions was assessed by forming 2×2 tables showing the relation between predictions and actual decisions and computing a coefficient of contingency for each table. Means and standard errors of the contingency coefficients are shown in Fig. 8. A repeated-measures ANOVA indicated significant effects of bandwidth $[F(5,30)=4.6; p<0.01]$ and bandwidth $\times$ predictor $[F(5,30)=4.9; p<0.01]$. The effect of predictor (weights versus model) was not significant $[F(1,6)=0.002; p=0.9]$. The interaction reflects the fact that perceptual weights are generally less accurate than the model at the narrow bandwidths, but are more accurate than the model at the widest bandwidth, where the weights provide a description of the individual differences not provided by the model.

The individual differences and low correlation between threshold and perceptual weight in the widest bandwidth condition suggests that the measurement of perceptual

weight may be of greatest value in understanding the contributions of individual components to the overall loudness of a complex sound under conditions in which the components are widely spaced in frequency and clearly audible. Under such conditions, an attentional component to loudness summation appears evident.

Studies in the literature are consistent with the idea that central auditory processes can influence loudness. Zwicker *et al.* (1957) observed increased variability in loudness matches with increasing bandwidth. In addition, they reported that different subjects used different criteria to judge the loudness of a complex sound, especially when the complex was widely spaced in frequency, but that individual subjects used consistent strategies. A second line of evidence that listening strategy influences loudness comes from data showing that cognitive factors influence loudness adaptation. Schlauch (1992) demonstrated that the loudness of adapting tones depended on whether or not subjects actively listened to the stimulus. Leibold and Werner (2001) demonstrated that loudness growth differed in conditions where subjects were required to listen to a broad versus narrow range of frequencies, providing additional evidence that central factors can influence loudness perception. In a related study of intensity processing, Doherty and Lutfi (1996) demonstrated considerable individual differences in the weighting functions of subjects in an intensity-discrimination task using multitone complexes. Therefore, it seems reasonable that subjects also adopt a variety of different listening strategies during tasks that measure the loudness of similar complex sounds.

V. SUMMARY AND CONCLUSIONS

(1) Consistent with Zwicker *et al.* (1957), loudness matches to a 1000-Hz tone indicate that loudness remained constant when all components were located within a single

critical bandwidth. Beyond the critical bandwidth, loudness increased with increasing frequency bandwidth at a rate of approximately 2.8 dB/octave.

(2) When individual components fell within the same critical band, all components received similar perceptual weight.

(3) The range of weights increased with increasing frequency separation, with increasing weight given to the lowest and highest frequency components.

(4) A relation between the perceptual weight and masked threshold of individual components was observed for all but the widest bandwidth condition.

(5) The loudness model proposed by Moore *et al.* (1997) provided a good account of the loudness matches, perceptual weights, and masked thresholds in all but the widest bandwidth condition.

(6) For complexes with widely spaced components, measures of perceptual weight provide information regarding the contribution of central auditory processes to loudness that existing models do not.

ACKNOWLEDGMENTS

Berg, B. G. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.

Bernstein, R. S., and Raab, D. H. (**1990**). "The effects of bandwidth on the detectability of narrow- and wideband signals," J. Acoust. Soc. Am. **88**, 2115–2125.

Buus, S., Müsch, H., and Florentine, M. (**1998**). "On loudness at threshold," J. Acoust. Soc. Am. **104**, 399–410.

Doherty, K. A., and Lutfi, R. A. (**1996**). "Spectral weights for overall level discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **99**, 1053–1058.

Fletcher, H., and Munson, W. A. (**1937**). "Relation between loudness and masking," J. Acoust. Soc. Am. **9**, 1–10.

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

Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2003**). "Asymmetry of masking between complex tones and noise: Partial loudness," J. Acoust. Soc. Am. **114**, 349–360.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hübner, R., and Ellermeier, W. (**1993**). "Additivity of loudness across critical bands: A critical test," Percept. Psychophys. **54**, 185–189.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," J. Acoust. Soc. Am. **28**, 85–88.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., and Nizam, A. (**1998**). "The correlation coefficient and straight-line regression analysis," in *Applied Regression Analysis and Other Multivariable Methods*, edited by A. Kugushev (Duxbury, Pacific Grove), pp. 88–103.

Kortekaas, R., Buus, S., and Florentine, M. (**2003**). "Perceptual weights in auditory level discrimination," J. Acoust. Soc. Am. **113**, 3306–3322.

Leibold, L. J., and Werner, L. A. (**2001**). "The effect of listening strategy on loudness growth in normal-hearing adults," (Association for Research in Otolaryngology), Abstract 22051.

Leibold, L. J., and Werner, L. A. (**2003**). "Listening strategies used by normal-hearing adults during loudness estimation," J. Acoust. Soc. Am. **112**, 2273.

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

Lutfi, R. A. (**1989**). "Informational processing of complex sound. I. Intensity discrimination," J. Acoust. Soc. Am. **86**, 934–944.

Lutfi, R. A. (**1995**). "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," J. Acoust. Soc. Am. **97**, 1333–1334.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness," J. Audio Eng. Soc. **45**, 224–237.

Richards, V. M., and Zhu, S. (**1994**). "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434.

Schaeffer, T. H., Gales, R. S., Shewmaker, C. A., and Tompson, P. O. (**1950**). "The frequency selectivity of the ear as determined by masking experiments," J. Acoust. Soc. Am. **22**, 490–496.

Scharf, B. (**1961**). "Loudness summation under masking," J. Acoust. Soc. Am. **33**, 503–511.

Scharf, B. (**1962**). "Loudness summation and spectral shape," J. Acoust. Soc. Am. **34**, 228–233.

Scharf, B. (**1964**). "Partial masking," Acustica **14**, 16–23.

Schlauch, R. S. (**1992**). "A cognitive influence on the loudness of tones that change continuously in level," J. Acoust. Soc. Am. **92**, 758–765.

Schooneveldt, G. P., and Moore, B. C. J. (**1989**). "Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration," J. Acoust. Soc. Am. **85**, 273–281.

Spiegel, M. F. (**1981**). "Thresholds for tones in maskers of various bandwidths and for signals of various bandwidths as a function of signal frequency," J. Acoust. Soc. Am. **69**, 791–795.

Stecker, G. C., and Hafter, E. (**2002**). "Temporal weighting in sound localization," J. Acoust. Soc. Am. **112**, 1046–1057.

Stellmack, M. A., Willihnganz, M. A., Wightman, F. L., and Lutfi, R. A. (**1997**). "Spectral weights in level discrimination by preschool children: Analytic listening conditions," J. Acoust. Soc. Am. **101**, 2811–2821.

Willihnganz, M. S., Stellmack, M. A., Lutfi, R. A., and Wightman, F. L. (**1997**). "Spectral weights in level discrimination by preschool children: Synthetic listening conditions," J. Acoust. Soc. Am. **101**, 2803–2810.

Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). "Critical bandwidth in loudness summation," J. Acoust. Soc. Am. **29**, 548–557.

The effects of low- and high-frequency suppressors on psychophysical estimates of basilar-membrane compression and gain[a]

Ifat Yasin[b]
Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, United Kingdom

Christopher J. Plack
Department of Psychology, Lancaster University, Lancaster, LA1 4YF, United Kingdom

Physiological studies suggest that the increase in suppression as a function of suppressor level is greater for a suppressor below than above the signal frequency. This study investigated the pattern of gain reduction underlying this increase in suppression. Temporal masking curves (TMCs) were obtained by measuring the level of a 2.2-kHz sinusoidal off-frequency masker or 4-kHz on-frequency sinusoidal masker required to mask a brief 4-kHz sinusoidal signal at 10 dB SL, for masker-signal intervals of 20–100 ms. TMCs were also obtained in the presence of a 3- or 4.75-kHz sinusoidal suppressor gated with the 4-kHz masker, for suppressor levels of 40–70 dB SPL. The decrease in gain (increase in suppression) as a function of suppressor level was greater with a 3-kHz suppressor than with a 4.75-kHz suppressor, in line with previous findings. Basilar membrane input-output (I/O) functions derived from the TMCs showed a shift to higher input (4-kHz masker) levels of the low-level (linear) portion of the I/O function with the addition of a suppressor, with partial linearization of the function, but no reduction in maximum compression. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713675]

I. INTRODUCTION

Physiological studies of auditory suppression show that the vibration of the basilar membrane (BM), or the firing rate of an auditory nerve fiber, in response to a signal at a given frequency, may be reduced if a second signal (suppressor) is simultaneously presented at a different frequency (e.g., Delgutte, 1990; Geisler and Nuttall, 1997). Suppression of one signal by another is probably a consequence of nonlinear cochlear processing involving an active mechanism (Cheatham and Dallos, 1990). This active mechanism applies level-dependent gain to the BM response to signals with frequencies at, or close to, the characteristic frequency (CF) of each place on the BM. The amount of gain applied to the BM response (output) is dependent on the signal level (input) such that the gain applied is less for high-level signals than for low-level signals. This leads to a compressive BM input-output (I/O) function with a slope of about 0.2 dB/dB for midlevel signals (Recio *et al.*, 1998). It is thought that suppression is caused by a reduction in cochlear gain.

Physiological measurements show that there is a greater increase in suppression with increasing suppressor level for suppressors lower, rather than higher in frequency, than the signal (Javel *et al.*, 1983; Costalupes *et al.*, 1987; Delgutte,

1990). Suppression can be inferred from the shift to higher signal levels of the neural rate-level function (a plot of neural discharge rate versus signal level) in the presence of a suppressor (e.g., Javel, 1981). Neural studies suggest that the increase in suppression with suppressor level is about 1–3 dB/dB for a suppressor lower in frequency than the signal, and about 0.15–0.7 dB/dB for a suppressor higher in frequency than the signal (Costalupes *et al.*, 1987; Delgutte, 1990). Similarly, suppression at the BM level can be inferred from the shift to higher signal levels of the BM I/O function (a plot of BM vibration versus input sound level) in the presence of a suppressor. In comparing neural versus BM suppression data, it has to be noted that whereas the total response of the signal plus suppressor is represented in the neural rate of firing, the BM response represents separately filtered components of the signal and suppressor, hence the total BM excitation at the signal place does not necessarily decrease in conditions in which the signal is suppressed. Both Cooper (1996) and Ruggero *et al.* (1992) showed a greater increase in suppression for a suppressor with a frequency lower than that of the signal (1 dB/dB) than for a suppressor with a frequency higher than that of the signal (0.3–0.5 dB/dB).

Psychophysical estimates of suppression can be obtained from a comparison of the forward-masked threshold of a pure-tone signal in the presence of a masker of the same frequency, with and without a suppressor at another frequency gated with the masker. The suppressor acts to reduce the response of the BM to the masker (suppressee) resulting

---

in a reduction of the masked threshold of the signal (Hout-gast, 1974). Alternatively, some studies have estimated the increase in suppression with suppressor level by measuring the pulsation threshold. For a pulsation-threshold measurement, a signal is alternated with a masker. The pulsation threshold can be defined as the lowest signal level at which the signal is heard as pulsing on and off, rather than being heard as continuous. Psychophysical studies investigating the increase of suppression with suppressor level using this technique (Houtgast, 1974; Duifhuis, 1980) have shown that the decrease in pulsation threshold with suppressor level is greater when a suppressor lower, rather than higher in frequency than the signal, is gated with the masker. For instance, for a 1-kHz signal and masker, the rate at which pulsation threshold decreases with increasing level of a 0.4-kHz suppressor is about 3 dB/dB, while the rate at which pulsation threshold decreases with increasing level of a 1.2-kHz suppressor is about 1 dB/dB (Duifhuis, 1980; Figs. 4 and 9). So, both physiological and psychophysical measures demonstrate differences in the growth of suppression with suppressor level between suppressors higher and lower in frequency than the signal. However, it remains unclear how this difference in suppression growth relates to the pattern of gain reduction in the BM response.

A suppressor may reduce the gain of the BM response by causing saturation of the outer hair cell (OHC) mechanical-to-electrical transduction process. This would lead to a reduction of the gain applied by the cochlear amplifier (Nobili and Mammano, 1996). Since compression is related to the variation in gain with level, suppression may also affect compression of the BM response. One way in which a suppressor may be hypothesized to affect the BM response is by producing a gain reduction that declines progressively with level, effectively linearizing the midportion of the BM I/O function and reducing BM compression (e.g., Ruggero et al., 1992). However, for suppressor levels substantially higher than the level of the masker, the gain of the cochlear amplifier may be determined mainly by the suppressor level (Pang and Guinan, 1997). For a constant suppressor level, this will result in a near linear growth in BM response with masker level, until the masker level becomes high enough to dominate the OHC receptor potential. For higher masker levels, the gain will vary with masker level, resulting in a compressive response. In other words, the hypothesis predicts that a suppressor reduces the gain at low masker levels but not at high levels, thereby shifting the I/O function to the right, leaving a residual region of compression at high levels.

Physiological studies of induced cochlear hearing loss are often used to investigate the effect of reduced OHC function on the pattern of gain applied to the BM response. Some of these studies lend support to the idea that a reduction in OHC function does not necessarily lead to a linearization of the midportion of the BM I/O response (Yates, 1990). These studies report a reduction in gain at low levels with minimal effect on maximum compression (chinchilla data: Ruggero and Rich, 1991; guinea-pig data: Murugasu and Russell, 1995).

Two recent psychophysical studies have examined suppressive gain reduction in normal-hearing listeners. Yasin and Plack (2003) investigated the effect of an increase in level (50–60 dB SPL) of a high-frequency suppressor on estimates of the BM response. A 4.75-kHz suppressor was presented simultaneously with a 4-kHz masker. Both the suppressor and masker were presented before a fixed-level 4-kHz signal. By measuring the increase in the 4-kHz masker required to maintain masked threshold, the effect of the suppressor on the compression and gain applied to the 4-kHz masker could be estimated. Both compression and gain estimates were found to decrease in the presence of a suppressor. Wojtczak and Viemeister (2005) showed similarly reduced estimates of compression and gain for a 4-kHz forward masker simultaneously presented with a 4.8-kHz suppressor.

The aim of the present study was to investigate the effect of increases in suppressor level on estimates of compression and gain, for suppressor frequencies below and above the signal frequency. Psychophysical BM I/O functions can be derived by measuring the level of a forward masker required to mask a fixed low-level signal for masker frequencies equal to, or about an octave below, the signal frequency (Nelson et al., 2001). Forward-masked thresholds are obtained as a function of masker-signal interval and plotted as a temporal masking curve (TMC). For a given masker-signal interval, it is assumed the BM response to the masker at the signal place is constant at threshold (regardless of masker frequency). It is further assumed that the BM response at that place to the low-frequency masker is linear, and the low-frequency masker threshold is therefore an estimate (give or take an additive constant in decibels) of the BM response needed to mask the signal. Hence, for a given masker-signal interval the level of the 2.2-kHz masker level (output level) is plotted against the level of the 4-kHz masker (input level). This is repeated for a range of masker-signal intervals to obtain an estimate of the BM I/O function at the signal place.

The present study extends the earlier findings of Yasin and Plack (2003) using the TMC technique, by deriving BM I/O response functions for a 4-kHz signal without a suppressor, and with increasing levels (40–70 dB SPL) of a 3- or 4.75-kHz suppressor. The derived I/O functions can then be used to quantify the increase in suppression as a function of suppressor level and the effect of the suppressor on BM gain and compression at the signal place. According to the above-described hypothesis, we would predict that the suppressor would produce a gain reduction at low masker levels but not at high masker levels, linearizing the response at low levels, but having little effect on the compression at high levels.

II. METHOD

A. Conditions

In the no-suppressor conditions, masker levels at threshold were obtained for a 2.2- or 4-kHz forward masker preceding a 4-kHz sinusoidal signal. In the suppressor conditions, masker levels at threshold were obtained for a 4-kHz forward masker presented simultaneously with a 3- or 4.75-kHz suppressor, both preceding a 4-kHz sinusoidal sig-

nal. Suppressor levels of 40, 50, 60, or 70 dB SPL were tested. The two no-suppressor and eight suppressor conditions were presented as a randomized set for masker-signal intervals of 20, 30, 40, 60, 80, and 100 ms.

B. Stimuli

The 4-kHz signal had 4-ms raised-cosine onset and offset ramps with a 0-ms steady-state portion. Both the masker and suppressor had 2-ms raised-cosine onset and offset ramps and 100-ms steady-state portions. The suppressor was gated with the masker. The signal was presented at 10 dB SL for all listeners. The masker-signal interval was specified as the silent time interval between the zero-points in the envelopes at the end of the masker and the beginning of the signal. A notched noise, with a passband spectrum level of 0 dB, was gated with the masker (and suppressor if present). The notched noise had a center frequency of 4 kHz, a notch width of 800 Hz (3-dB downpoints), and attenuation slopes of 90 dB/octave. The notched noise acted as a temporal cue to reduce confusion (Neff, 1986) particularly when the signal frequency was the same as that of the signal (4 kHz).

All stimuli were digitally generated by a PC using a sampling rate of 48 kHz and output via a soundcard with a 24-bit resolution. Antialiasing was provided by built-in filters. The stimuli were presented to the right channel of a set of Sennheiser HD 580 headphones. The headphone input came directly from the output of the soundcard DAC.

C. Procedure

Listeners were tested individually while seated in an IAC double-walled sound-attenuating booth. Listeners observed a computer screen through a window in the sound booth. A block of trials began with the presentation of a light on a computer-simulated response box presented on the computer screen. Listeners started a block of trials by pressing a start key on a keyboard inside the sound booth. The experiment used a two-interval, two-alternative forced choice adaptive tracking procedure. The signal level was fixed at 10 dB SL and the masker level was varied adaptively to obtain the masker level required to achieve 70.7% correct (Levitt, 1971).

The length of each observation interval was indicated by a light on the response box. On each trial, the masker (and suppressor, if present), was present in both intervals, the signal was presented at random in one of the intervals. The interstimulus interval was 500 ms. The task within a trial was to select the signal interval. Listeners responded by pressing the appropriate response key. After a response, visual feedback was provided by the presentation of a colored light. The masker level was decreased after an incorrect response and increased after every two consecutive correct responses. A reversal was counted every time the masker level changed direction. For the first four reversals the masker level was varied in steps of 4 dB. For the following 12 reversals the steps were 2 dB. For each block of trials the value of threshold was estimated by averaging the masker levels for the last 12 reversals. If the standard deviation of the estimated threshold value was greater than 7 dB, the block of

trials was repeated. Data collection commenced after 20 h of practice for each listener. Estimates of threshold were obtained until performance was stable; a post-hoc stable estimate was taken to be a standard deviation of less than 8 dB for three successive estimates of threshold. The mean threshold was calculated from these three estimates of threshold. The maximum masker level produced by the system was 102 dB SPL. The system peak clipped at this level, producing distortion at the headphone output. If a masker level of 102 dB SPL was reached more than twice within a block of trials, the estimate of threshold from that block was discarded.

D. Listeners

Five normal-hearing listeners were tested, CO, ES, IY, LS, and PP. All, except IY, were paid for their services. Absolute thresholds were obtained for signal frequencies of 0.25, 0.5, 1, 2, 4, and 8 kHz, using an adaptive track in which the signal level (200-ms steady state, 10-ms raised-cosine onset and offset ramps) was varied in a two-down, one-up procedure. All listeners had absolute thresholds better than 20 dB HL (ANSI, 1996). Absolute thresholds for the 4-kHz signal used in the experiment, for CO, ES, IY, LS, and PP, were 9.0, 9.5, 13.5, 9.5, and 4.8 dB SPL, respectively.

III. RESULTS

Listener ES failed to reach a stable threshold level, with the standard deviation of successive threshold values differing by much more than 8 dB. For this reason, this listener's data were omitted from the subsequent analyses. Data for some listeners could not be obtained for the longer masker-signal intervals (CO; 80–100 ms, PP; 90–100 ms); the masker levels consistently reached levels of 102 dB SPL producing distortion at the headphone output. For a suppressor level of 70 dB SPL, the thresholds were much lower than the thresholds obtained without a suppressor; both the 3- and 4.75-kHz suppressors were acting as additional maskers. For this reason, the forward-masked thresholds obtained with a suppressor level of 70 dB SPL have not been included in the analyses.

A. Derived estimates of compression, gain, and suppression

Individual TMCs for each of the subjects, CO, IY, LS, and PP for masker-signal intervals of 20–70 ms are presented in Fig. 1. TMCs are shown for each listener when no suppressor is present (closed symbols) and when a 3- or 4.75-kHz suppressor (open symbols; left and right column, respectively) is presented with the 4-kHz masker. The level of a 2.2- or 4-kHz masker required to mask the signal is shown as a function of masker-signal interval. The mean TMCs for listeners CO, IY, LS, and PP are presented in Fig. 2. The mean standard error for the no-suppressor condition for a 2.2- or 4-kHz masker was 1.98 and 2.24 dB, respectively. The mean standard error when the 4-kHz masker was presented with a 3- or 4.75-kHz suppressor was 2.27 and 1.98 dB, respectively. In both Figs. 1 and 2 only standard errors larger than the symbol size are shown.

I. Yasin and C. J. Place: Effect of suppression on gain and compression
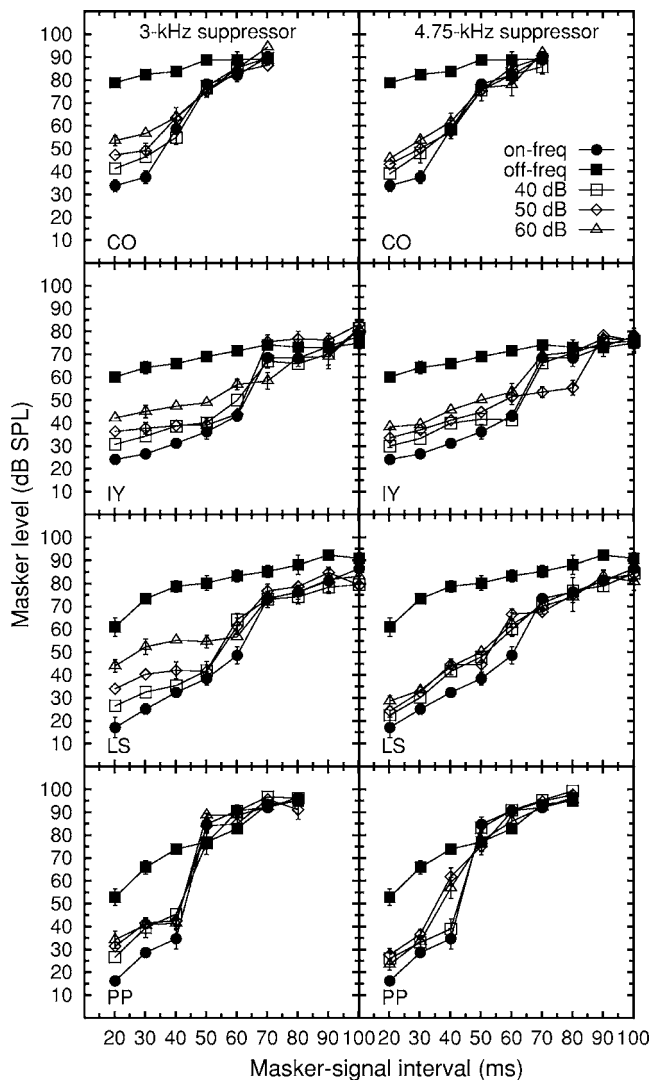
FIG. 1. Individual TMCs, showing masker level at threshold as a function of masker-signal interval. Each row of panels presents the TMCs for each of the subjects CO, IY, LS, and PP. The left column shows TMCs when a 3-kHz suppressor was gated with the 4-kHz masker. The right column shows TMCs when a 4.75-kHz suppressor was gated with the 4-kHz masker. On-frequency (4 kHz) and off-frequency (2.2 kHz) TMCs are represented by closed circles and squares, respectively. Suppressor levels of 40, 50, and 60 dB SPL are represented by open squares, diamonds, and triangles, respectively. Only error bars larger than the symbol size are shown.

For the no-suppressor condition, the increase in the 4-kHz masker level is shallow for masker-signal intervals below 40 ms. For masker-signal intervals longer than 40 ms the 4-kHz masker level increases steeply with increasing interval. In comparison, the increase in the 2.2-kHz masker level is relatively shallow for masker-signal intervals between 20 and 70 ms. The steep increase of the 4-kHz masker level relative to the shallow increase of the 2.2-kHz masker is thought to reflect frequency-dependent compression. The BM response is compressive for frequencies close to CF (leading to a steep TMC) and linear for frequencies much lower than CF. Hence, the same increase in BM response requires a greater increase in level for the 4-kHz masker level than for the 2.2-kHz masker. The shapes of the TMCs are similar under suppression. However, the 3- and 4.75-kHz suppressors produce an increase in the level of the 4-kHz
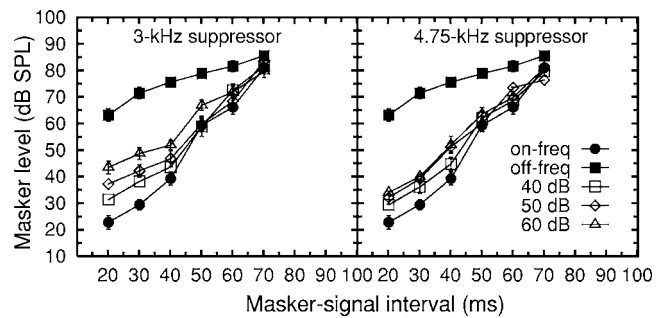
FIG. 2. Mean TMCs, showing masker level at threshold as a function of masker-signal interval. The left panel shows TMCs when a 3-kHz suppressor was gated with the 4-kHz masker. The right panel shows TMCs when a 4.75-kHz suppressor was gated with the 4-kHz masker. On-frequency (4 kHz) and off-frequency (2.2 kHz) TMCs are represented by the closed circles and squares, respectively. Data for suppressor levels of 40, 50, and 60 dB SPL are represented by open squares, diamonds, and triangles, respectively. Only error bars larger than the symbol size are shown.

masker for masker-signal intervals less than about 50 ms. For a given increase in suppressor level, the increase in the 4-kHz masker level is generally greater with the addition of a 3-kHz suppressor than with the addition of a 4.75-kHz suppressor.

BM I/O response functions can be derived from the TMC data by assuming that the BM response for the 2.2-kHz masker grows linearly with increases in masker level. If the 2.2-kHz masker level is plotted against the 4-kHz masker level, paired by masker-signal interval, then the function describes the BM response to the 4-kHz masker at the signal place. A similar approach has been used previously for psychophysical data (Nelson *et al.*, 2001; Plack and Drga, 2003; Yasin and Plack, 2003; Rosengard *et al.*, 2005).

The BM I/O functions derived from the mean data are shown in Fig. 3. Each function presents the threshold levels for the 2.2-kHz off-frequency masker (output level), versus the 4-kHz on-frequency masker (input level), paired according to the masker-signal interval (20, 30, 40, 50 60, and 70 ms). The left and right panels present derived I/O functions when a 3- or 4.75-kHz suppressor is present (open sym-
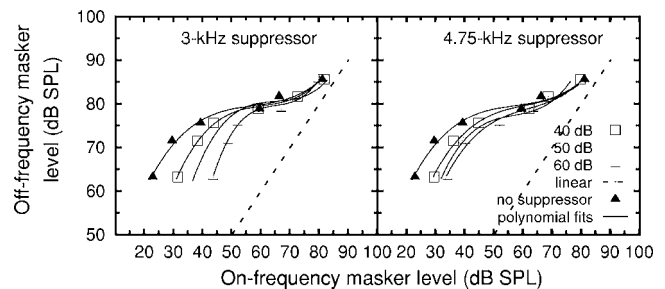
FIG. 3. Mean BM input-output functions. Each function represents the level of the 2.2-kHz off-frequency masker at threshold (output level) vs the level of the 4-kHz on-frequency masker at threshold (input level), paired for masker-signal intervals of 20, 30, 40, 50, 60, and 70 ms. The left and right panels show input-output functions when a 3- and 4.75-kHz suppressor is gated with the 4-kHz masker, respectively. Closed triangles represent the no-suppressor condition. Data for suppressor levels of 40, 50, and 60 dB SPL are represented by open squares, diamonds, and triangles, respectively. Solid lines represent third-order polynomial functions fit to the experimental data. The dashed line represents a linear I/O function. Error bars are omitted for clarity.
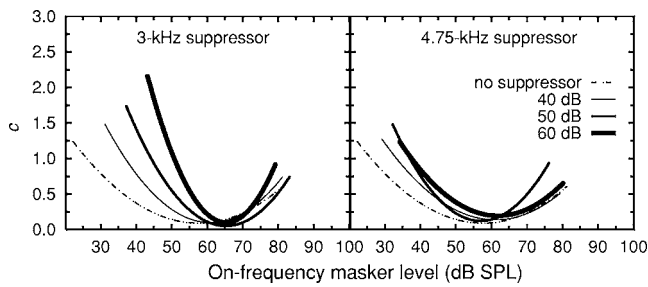
FIG. 4. Mean values of c (from the first derivative of the polynomial functions fit to the input-output data in Fig. 3) as a function of the 4-kHz on-frequency masker level. The left and right panels show the values of c obtained when a 3- or 4.75-kHz suppressor was gated with the 4-kHz masker, respectively. Lines of increasing thickness represent increasing levels of the suppressor. The dashed line represents the no-suppressor condition.

FIG. 5. Mean values of gain as a function of the 4-kHz on-frequency masker level. Gain was estimated from the polynomial fits to the input-output functions of Fig. 3 as the difference between the off-frequency masker level estimated from the polynomial fits to the I/O functions and the off-frequency masker levels for the case when no gain is applied (i.e., when on-frequency masker level equals off-frequency masker level). The left and right panels show the values of gain obtained when a 3- or 4.75-kHz suppressor was gated with the 4-kHz masker, respectively. Lines of increasing thickness represent increasing levels of the suppressor. The dashed line represents the no-suppressor condition.

bols). The I/O functions for the no-suppressor condition are shown by the closed symbols. For the no-suppressor condition and for on-frequency masker levels up to 40 dB SPL, the slope of the I/O function is close to 1. For on-frequency masker levels of 40–70 dB SPL, the slope of the no-suppressor I/O function is less than 1; there is a greater increase in the on-frequency masker level required per increase in the off-frequency masker level. The addition of a 3- or 4.75-kHz suppressor produces a shift to the right (higher on-frequency masker levels) of the low-level, linear, portion of the I/O function. For a given increase in suppressor level, this shift is greater with a 3- than with a 4.75-kHz suppressor.

For a given on-frequency masker level (input level), the slope of the I/O function provides an estimate of the compression exponent c (the compression applied to the BM response is given by $1/c$). To find the slope, the mean I/O functions of Fig. 3 were fitted with third-order polynomial functions. The polynomial fits to the I/O functions are shown by the solid lines in Fig. 3. For each condition, the value of c was estimated by taking the first derivative of the polynomial fits to the I/O functions. The derived values of c plotted as a function of the 4-kHz on-frequency masker level (input level) are shown in Fig. 4. Each panel of Fig. 4 presents four compression functions, with increasing line thickness representing increasing suppressor level (40–60 dB SPL). The dashed line represents the compression function obtained in the absence of a suppressor. As on-frequency masker level increases from 30 to 60 dB SPL, c decreases (compression increases). The minimum value of c (maximum compression) for the no-suppressor condition was 0.09 at an on-frequency masker level of 57 dB SPL. The average value of c was 0.3 for the unsuppressed I/O function for an on-frequency masker level range of 30–80 dB SPL [this range of on-frequency masker levels is similar to the range used to estimate c for a 4-kHz signal by Yasin and Plack (2003)]. The minimum values of c with a 3-kHz suppressor at levels of 40, 50, and 60 dB SPL were 0.1, 0.05, and 0.1 at on-frequency masker levels of 60, 66, and 65 dB SPL. The minimum values of c with a 4.75-kHz suppressor at levels of 40, 50, and 60 dB SPL were 0.1, 0.1, and 0.2 at on-frequency masker levels of 61, 57, and 62 dB SPL. As on-frequency masker levels increase to about 80 dB SPL, c increases again

to a value of about 0.75. For on-frequency masker levels between 30 and 60 dB SPL and for successive suppressor level increases from 40 to 60 dB SPL, c grows more quickly with increasing suppressor level for a 3- than for a 4.75-kHz suppressor. For low on-frequency masker levels, the value of c exceeds 1 for both suppressor frequencies and for all suppressor levels, suggesting an expansive nonlinearity.

The change in gain with on-frequency masker level was estimated from the polynomial fits to the I/O functions of Fig. 3. For a given on-frequency masker level, gain was estimated as the difference between the off-frequency masker level estimated from the polynomial fits to the I/O functions and the off-frequency masker level predicted for the case when no gain is applied (i.e., when on-frequency masker level equals off-frequency masker level; $y=x$). Figure 5 shows the variation in gain as a function of on-frequency masker level with either a 3-kHz suppressor (left panel) or 4.75-kHz suppressor (right panel) for increasing suppressor levels of 40, 50, or 60 dB SPL (shown by increasing line thickness). The dashed line represents the gain in the absence of a suppressor. At lower on-frequency masker levels the gain initially increases with masker level. This is most apparent in the data for a 3-kHz suppressor and reflects the expansive initial portion of the I/O curve (Fig. 4). Although this effect may partly be an artifact of the curve-fitting procedure, there does appear to be a steeper than linear growth for the lowest on-frequency masker levels in the raw data. After the maximum at lower levels, gain steadily decreases as on-frequency masker level increases. The presence of a suppressor causes a reduction in gain at low-to-medium masker levels. There is a greater decrease in gain with increasing suppressor level for a 3-kHz suppressor than for a 4.75-kHz suppressor.

B. The increase of suppression with suppressor level

The amount of suppression as a function of on-frequency masker level was estimated as the difference between the off-frequency masker levels generated by the polynomial fits to the I/O functions in Fig. 3 with and without a

I. Yasin and C. J. Plack: Effect of suppression on gain and compression
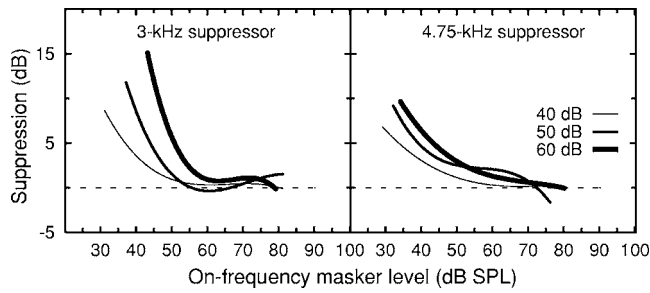
FIG. 6. Mean amount of suppression as a function of the 4-kHz on-frequency masker level. The amount of suppression was calculated as the difference between the 2.2-kHz off-frequency masker level at masked threshold, with and without a suppressor gated with the 4-kHz on-frequency masker. The values of the off-frequency masker were generated from the third-order polynomial fits to the I/O functions of Fig. 3, as described in the text. The left and right panels show the amount of suppression when a 3- or 4.75-kHz suppressor was gated with the 4-kHz masker, respectively. Lines of increasing thickness represent increasing levels of the suppressor. The dashed line represents zero suppression.

suppressor. This estimate reflects the decrease in the BM response, rather than the equivalent decrease in physical masker level (difference between the masker levels, with and without a suppressor); the latter can be much greater than the former because of the compression. [A similar estimate of suppression was made by Yasin and Plack (2005).] The amount of suppression was calculated for suppressor levels of 40, 50, and 60 dB SPL. Figure 6 shows the mean amount of suppression as a function of on-frequency masker level. The left and right panels show the amount of suppression with a 3- and 4.75-kHz suppressor, respectively. Lines of increasing thickness represent increasing suppressor levels. For on-frequency masker levels of 40–60 dB SPL, the amount of suppression is greater for a 3-kHz suppressor than for a 4.75-kHz suppressor. The maximum amount of suppression with a 3-kHz suppressor was 8, 11, and 15 dB for suppressor levels of 40, 50, and 60 dB SPL. The maximum amount of suppression with a 4.75-kHz suppressor was 6, 9, and 9 dB for the same suppressor levels. For on-frequency masker levels of 50–60 dB SPL, the amount of suppression is similar for a 3- and 4.75-kHz suppressor. For higher on-

frequency masker levels the amount of suppression diminishes to zero for both suppressor frequencies. Negative values of suppression for the highest on-frequency masker levels arise for cases in which the off-frequency masker level without a suppressor is lower than the off-frequency masker level with a suppressor.

For a given on-frequency masker level, there is a greater increase in the amount of suppression with increase in suppressor level for a 3-kHz suppressor than for a 4-kHz suppressor. If the amount of suppression is plotted as a function of the level of a 3- or 4.75-kHz suppressor, the slope of this function provides an estimate of the rate of increase in suppression with suppressor level. The values of these slopes are presented in Table I. Comparing just the positive values of slopes for a 3-kHz suppressor, the increase of suppression with suppressor level is greater (higher slope value) for an increase in suppressor level from 50 to 60 dB SPL than for an increase in suppressor level from 40 to 50 dB SPL. The reverse appears to be true for a 4.75-kHz suppressor; there is a trend for the increase in suppression with suppressor level to be slightly greater for an increase in suppressor level from 40 to 50 dB SPL than for an increase in suppressor level from 50 to 60 dB SPL. Overall, addition of a 3-kHz suppressor results in a greater increase in suppression for a given suppressor-level increase than that produced by a 4.75-kHz suppressor.

C. Interaction between suppression and compression

In order to examine the relationship between the increase in suppression and the estimate of compression, c is plotted as a function of the amount of suppression in Fig. 7, paired by on-frequency masker level. This gives a tick-like function, with pairings for the lowest on-frequency masker levels at the tip of the long-arm of the tick, with on-frequency masker level increasing down the long arm of the tick. For the 3-kHz suppressor condition (left panel) the value of c and the corresponding amount of suppression are paired by on-frequency masker levels in the ranges of 31–81, 37–81, and 43–79 dB SPL, for suppressor levels of 40, 50,

TABLE I. The rate of increase in suppression for the 4-kHz on-frequency masker level (input level) between 32 and 60 dB SPL. The value of the slope (dB/dB) provides an estimate of suppression as a function of increase in suppressor level (from 40 to 50 dB SPL, or 50 to 60 dB SPL), for a given input level and suppressor frequency.

	Suppressor level increase for 40 to 50 dB SPL			Suppressor level increase from 50 to 60 dB SPL	
	3-kHz suppressor	4.75-kHz suppressor		3-kHz suppressor	4.75-kHz suppressor
Input level dB SPL	Slope (dB/dB)	Slope (dB/dB)	Input level dB SPL	Slope (dB/dB)	Slope (dB/dB)
32		0.36	32.00		
36		0.23	36.00		0.20
40	0.49	0.16	40.00		0.18
44	0.27	0.13	44.00	0.85	0.13
48	0.11	0.13	48.00	0.52	0.07
52	0.01	0.15	52.00	0.31	0.01
56	−0.05	0.17	56.00	0.19	−0.04
60	−0.07	0.18	60.00	0.13	−0.08
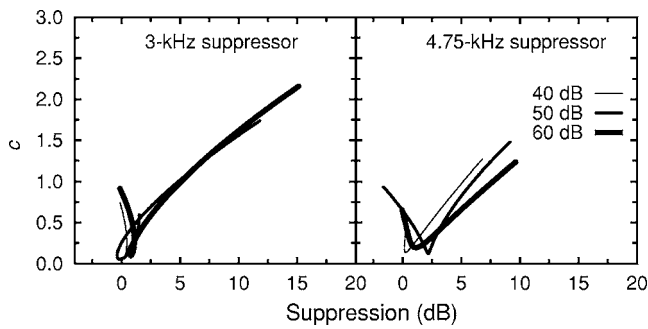
FIG. 7. The compression exponent *c*, as a function of the amount of suppression, paired by the 4-kHz on-frequency masker level. For the 3-kHz suppressor condition (left panel), *c* and the amount of suppression are paired by on-frequency masker levels in the ranges of 31–81, 37–81, and 43–79 dB SPL, for suppressor levels of 40, 50, and 60 dB SPL, respectively. For the 4.75-kHz suppressor condition (right panel) *c* and the amount of suppression are paired by on-frequency masker levels in the ranges of 29–79, 32–76, and 34–80 dB SPL for suppressor levels of 40, 50, and 60 dB SPL, respectively. Lines of increasing thickness represent increasing level of the suppressor. The pairings for the lowest on-frequency masker levels are at the tip of the long arm of the tick-like function, with on-frequency masker level increasing down the long arm of the tick.

and 60 dB SPL, respectively. Likewise, for the 4.75-kHz suppressor, condition (right panel) *c* and the amount of suppression are paired by on-frequency masker levels in the ranges of 29–79, 32–76, and 34–80 dB SPL for suppressor levels of 40, 50, and 60 dB SPL, respectively. As on-frequency masker level increases, suppression continues to decrease while compression increases. Both panels show that maximum compression is approximately the same for both suppressor frequencies and for all suppressor levels. For on-frequency masker levels higher than this (short arm of the tick) there is a sharp increase in *c* (decrease in compression), while there is relatively little change in the amount of suppression. For a 3-kHz suppressor the relationship between compression and suppression appears to be almost unaffected by suppressor level. That is, for a given amount of suppression, the compression exponent remains relatively constant. This is to some extent also true for a 4.75-kHz suppressor although a slight increase in compression (decrease in *c*) is evident for increasing suppressor levels. This finding represents the relatively smaller shift of the I/O function with a 4.75-kHz suppressor (see Fig. 3), such that increasing input levels correspond to more compressive regions of the derived BM I/O functions as the suppressor level is increased from 40 to 60 dB SPL.

IV. DISCUSSION

A. The effects of suppressor frequency

The amount of suppression was estimated as the difference between the off-frequency masker levels (2.2-kHz masker levels) of the derived BM I/O functions with and without a suppressor. For suppressor levels of 40, 50, and 60 dB SPL the maximum amount of suppression is greater with a 3-kHz suppressor (8, 11, and 15 dB, respectively) than with a 4.75-kHz suppressor (6, 9, and 9 dB, respectively). This amount of suppression is within the range reported by some recent studies [8–10 dB: Yasin and Plack

(2003); 5–10 dB: Wojtczak and Viemeister (2005)] for a 4-kHz signal, 4-kHz masker, and a high-frequency suppressor. A comparison of simultaneous and nonsimultaneous masking of a 4-kHz signal by a 4- or 2.4-kHz sinusoidal or narrow-band masker produced comparable suppression estimates for signal levels close to 10 dB SL (Oxenham and Plack, 1998; Yasin and Plack, 2005). The range of on-frequency masker levels (30–50 dB SPL) for which suppression is maximal is also close to that found in previous studies of suppression [40–60 dB SPL: Oxenham and Plack (1998); 35–60 dB SPL: Yasin and Plack (2005)]. The current estimates of the amount of suppression are, however, lower than estimates from some earlier studies [e.g., 12–35 dB: Duifhuis (1980); 20–30 dB: Shannon (1986)]. In these studies a suppressor was presented simultaneously with a forward masker to mask a subsequently presented signal, and the signal level was varied to find threshold. Suppression was calculated as the decrease in signal threshold with a suppressor. Since the signal level was varied, the estimate of suppression may be an over-estimate of the reduction in the BM response, due to compression of the signal at mid levels.

The increase in suppression with suppressor level was greater for a 3-kHz suppressor than for a 4.75-kHz suppressor. For example, measured in terms of the decrease in the off-frequency masker level for an on-frequency masker level of 44 dB SPL, the increase of suppression with suppressor level for a 3-kHz suppressor was 0.27 dB/dB for a suppressor level increase from 40 to 50 dB SPL, and 0.85 dB/dB for a suppressor level increase from 50 to 60 dB SPL. The increase in suppression with suppressor level was noticeably less with a 4.75-kHz suppressor; 0.13 and 0.13 dB/dB for the same suppressor-level increases. These estimates of the increase in suppression with suppressor level are less than those reported in earlier psychophysical studies such as Duifhuis (1980). In this study, suppression was estimated as the difference in the input level; the decrease in signal threshold when a suppressor was simultaneously presented with a forward masker in a pulsation-threshold paradigm. Since there is the possibility that the signal itself may have been compressed at midlevels, these values may overestimate suppression of the BM response, and overestimate the increase in suppression of the BM response with suppressor level. The present estimates of the increase in suppression with suppressor level are also less than those observed in some physiological data [Neural: Javel (1981); Costalupes et al. (1987); BM: Ruggero et al. (1992); Rhode and Recio (2001)] which similarly estimated suppression as either a shift of the neural rate-level function to higher signal levels [equivalent to the horizontal shift of the derived BM I/O function to higher on-frequency masker levels (input) levels] or in the case of BM vibration measures, a shift of the BM I/O function to higher signal levels. Overall, the increase in suppression is greater for a suppressor of a frequency lower, rather than higher than that of the signal, which is in accordance with previous reports.

B. The effect of a suppressor on compression and gain estimates

The estimated mean c value of 0.3 from the polynomial fits to the I/O functions for the no-suppressor condition is close to the values obtained in previous psychophysical studies [$c=0.05-0.37$: Plack and Drga (2003); $c=0.3$: Yasin and Plack (2003); $c=0.26$: Yasin and Plack (2005)], but slightly higher (less compressive) than reported by other psychophysical and physiological studies [$c=0.2$: Yates *et al.* (1990); $c=0.13$: Ruggero *et al.* (1992); $c=0.16$: Oxenham and Plack (1997); $c=0.26$: Nelson *et al.* (2001); $c=0.2$: Plack *et al.* (2004).] However, a straightforward comparison of estimates of c across studies is difficult for the reason that the compression is estimated over a different range of levels, and for different signal frequencies. For instance, while Plack and Drga (2003) and Nelson *et al.* (2001) quote the maximum compression values estimated from the first derivative of the third-order polynomial functions fit to the derived BM I/O functions, Yasin and Plack (2003) estimate c from the midportion of a three-section fit to the derived BM I/O function. Furthermore, estimates of compression have been compared for a range of signal frequencies: 4-kHz signal (Plack and Drga, 2003; Yasin and Plack, 2003; Plack *et al.*, 2004; Yasin and Plack, 2005), 6-kHz signal (Oxenham and Plack, 1997; Nelson *et al.*, 2001) and 8-kHz signal (Ruggero *et al.*, 1992). In the present study, maximum compression without a suppressor occurs for an on-frequency masker (input) level of 57 dB SPL, corresponding well with other psychophysical studies (Nelson *et al.*, 2001; Nelson and Schroder, 2004; Yasin and Plack, 2005).

Curiously, in some cases the value of c for low on-frequency masker levels exceeded 1, indicating an expansive nonlinearity; this was especially obvious for a 3-kHz suppressor. An expansive nonlinearity has previously been reported in the presence of suppression by Yasin and Plack (2005). Although an expansive nonlinearity in the *absence* of suppression has been reported in some psychophysical (Nelson and Schroder, 2004) and physiological studies (Zinn *et al.*, 2000; Cooper and Rhode, 1992, 1995), there appears to be less evidence of an expansive nonlinearity in the presence of suppression. However, the physiological mechanism(s) underlying these expansive responses (in either the absence or presence of suppression) remains unclear. The expansion observed here may partly be a consequence of limited data points and data variability at low levels. It is also possible that the expansion reflects a limitation in the TMC technique at low masker levels, rather than expansion in the BM response.

The present results suggest that the decrease in gain (amount of suppression) produced by the suppressor is greatest at low on-frequency masker levels and becomes progressively less as level is increased, reaching zero for on-frequency masker levels close to the suppressor level (Fig. 6). Hence the effect of suppression is to shift the low-level portion of the BM I/O function to higher input levels, partially linearizing the response at medium levels, while leaving the high-level portions of the function unaffected (Fig. 3). The results are broadly consistent with the hypothesis described in Sec. I (Pang and Guinan, 1997), in which the gain of the cochlear amplifier is assumed to be determined by the fixed suppressor level for low masker levels, leading to a nearly linear response growth. The shift to higher on-frequency masker levels was greater with an increase in level of a 3-kHz suppressor than a 4.75-kHz suppressor. Furthermore, the present findings show that although the on-frequency masker level for which compression was maximal increased slightly in the presence of a 3- or 4.75-kHz suppressor, the value of maximal compression was almost unaffected by suppression, or by changes in suppressor level or frequency. A similar partial linearization and shift of the derived BM I/O function to higher input levels with the addition of a suppressor has previously been reported by Yasin and Plack (2003).

Some physiological studies show a constant shift of neural rate-level functions to higher signal levels with increasing suppressor level, with minimal changes in the slope of the function for a given firing rate (Sachs, 1969; Abbas and Sachs, 1976; Abbas, 1978; Javel *et al.*, 1978). This means that to maintain an equivalent rate of neural firing in the presence of a suppressor, the signal level for the fiber at CF is increased, while the slope of the function remains unaffected. A similar horizontal shift of the BM I/O function with addition of increasing levels of a suppressor is seen in some BM vibration studies (e.g., Rhode and Recio, 2001). These physiological findings are consistent with the present findings of a horizontal shift of the derived BM I/O function with the addition of a suppressor, with a small linearization of the I/O function (Ruggero *et al.*, 1992). The results are also consistent with the low-level attenuation model depicted in Plack *et al.* (2004; Fig. 7, panel C) that describes the effects of mild-to-moderate sensorineural hearing loss on the BM response. Some forms of hearing dysfunction and suppression seem to produce a similar shift in the I/O function.

C. The implication for the upward spread of masking

The greater increase in suppression with suppressor level for suppressors below the signal frequency than above the signal frequency is thought to play a role in the upward spread of masking (USM). The USM describes the situation when, for a given increase in masker level, a masker lower in frequency than the signal produces a greater increase in signal level at threshold than a masker higher in frequency than the signal (Wegel and Lane, 1924; Egan and Hake, 1950). This nonlinear growth of simultaneous masking may result from a combination of both an upward spread of suppressive and excitatory masking (Delgutte, 1990). Suppressive masking is thought to be the main contributor toward the growth of USM for relatively low-level signals (Delgutte, 1990). However, for midlevel signals suppression may actually reduce the extent of the nonlinear upward spread of excitatory masking by increasing the slope of the BM I/O function (Robles *et al.*, 1987; Ruggero *et al.*, 1992) decreasing the compression of the signal response. In contrast, other studies of the USM show that suppression by a masker that varies with the signal level may occur without changes to the slope of the derived BM I/O function [e.g., Yasin and Plack (2005)]. This means that the gain of the cochlear amplifier is

reduced leaving the slope of the derived BM I/O function relatively unaffected. The present findings lend further support to this idea. Suppression of midlevel inputs by a high-frequency suppressor appears to be mainly due to a reduction of gain at low levels, without a substantial decrease in maximum compression at higher levels.

V. CONCLUSIONS

(i) Suppression causes a horizontal shift to higher on-frequency masker (input) levels of the low-level (linear) portion of the BM I/O function, while leaving the response at input levels higher than the suppressor level almost unaffected.

(ii) The decrease in gain (increase in suppression) is greater for a 3-kHz suppressor than for a 4.75-kHz suppressor. Also, the rate of decrease of gain with increases in suppressor level is greater for a 3- than for a 4.75-kHz suppressor, in line with previous findings.

(iii) A given amount of suppression is associated with a nearly constant amount of compression in the I/O function, almost irrespective of suppressor level or suppressor frequency. Furthermore, although maximum compression occurs at a slightly higher input level with the addition of a 3- or 4.75-kHz suppressor, the *value* of maximum compression is almost unaffected.

Abbas, P. J. (**1978**). "Effects of stimulus frequency on two-tone suppression," J. Acoust. Soc. Am. **63**, 1878–1886.

Abbas, P. J., and Sachs, M. B. (**1976**). "Two-tone suppression in auditory-nerve fibers: Extension of a stimulus-response relationship," J. Acoust. Soc. Am. **59**, 112–122.

ANSI (**1996**). ANSI S3.6–1996, "Specification for Audiometers" (American National Standards Institute, New York).

Cheatham, M. A., and Dallos, P. (**1990**). "Comparison of low- and high-side two-tone suppression in inner hair cell and organ of Corti responses," Hear. Res. **50**, 193–210.

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

Cooper, N. P., and Rhode, W. S. (**1992**). "Basilar membrane mechanics in the hook region of cat and guinea pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts," Hear. Res. **63**, 163–190.

Cooper, N. P., and Rhode, W. S. (**1995**). "Nonlinear mechanics at the apex of the guinea pig cochlea," Hear. Res. **82**, 225–243.

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

Delgutte, B. (**1990**). "Two-tone rate suppression in auditory-nerve fibres: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Duifhuis, H. (**1980**). "Level effects in psychophysical two-tone suppression," J. Acoust. Soc. Am. **67**, 914–927.

Egan, J. P., and Hake, H. W. (**1950**). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 622–630.

Geisler, C. D., and Nuttall, A. L. (**1997**). "Two tone suppression of basilar-membrane vibrations in the base of the guinea pig cochlea using "low-side" suppressors," J. Acoust. Soc. Am. **45**, 986–999.

Houtgast, T. (**1974**). "Lateral suppression in hearing," Ph.D. thesis, Free University of Amsterdam, Amsterdam.

Javel, E. (**1981**). "Suppression of auditory nerve responses. I. Temporal analysis, intensity effects and suppression contours," J. Acoust. Soc. Am. **69**, 1735–1745.

Javel, E., Geisler, C. D., and Ravindra, A. (**1978**). "Two tone suppression in auditory nerve of the cat: Rate-intensity and temporal analysis," J. Acoust. Soc. Am. **63** 1093–1104.

Javel, E., McGee, J., Walsh, E. J., Farley, G. R., and Gorga, M. P. (**1983**). "Suppression of auditory nerve responses. II. Suppression threshold and growth, iso-suppression contours," J. Acoust. Soc. Am. **74**, 801–813.

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

Murugasu, E., and Russell, I. J. (**1995**). "Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea," Aud. Neurosci. **1**, 139–150.

Neff, D. L. (**1986**). "Confusion effects with sinusoidal and narrow-band forward maskers," J. Acoust. Soc. Am. **79**, 1519–1528.

Nelson, D. A., and Schroder, A. C. (**2004**). "Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2221–2233.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing, and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.

Nobili, R., and Mammano, F. (**1996**). "Biophysics of the cochlea. II. Stationary nonlinear phenomenology," J. Acoust. Soc. Am. **99**, 2244–2255.

Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioural measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.

Oxenham, A. J., and Plack, C. J. (**1998**). "Suppression and the upward spread of masking," J. Acoust. Soc. Am. **104**, 3500–3510.

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

Plack, C. J., and Drga, V. (**2003**). "Psychophysical evidence for auditory compression at low characteristic frequencies," J. Acoust. Soc. Am. **113**, 1574–1586.

Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.

Rhode, W. S., and Recio, A. (**2001**). "Multicomponent stimulus interactions observed in basilar membrane vibration in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **110**, 3140–3154.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1987**). "Two-tone suppression in the basilar membrane of the chinchilla," in *Neuroscience. The 2nd World Congress of Neuroscience (IBRO)* (Pergamon, Oxford).

Rosengard, P. S., Oxenham, A. J., and Braida, L. D. (**2005**). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **117**, 3028–3041.

Ruggero, M. A., and Rich, N. C. (**1991**). "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon basilar membrane," J. Neurosci. **11**, 1057–1067.

Ruggero, M. A., Robles, L., and Rich, N. C. (**1992**). "Two-tone suppression in the BM of the cochlea: Mechanical basis of auditory-nerve rate suppression," J. Neurophysiol. **68**, 1087–1099.

Sachs, M. B. (**1969**). "Stimulus-response relation for auditory-nerve fibers: Two-tone stimuli," J. Acoust. Soc. Am. **45**, 1025–1036.

Shannon, R. V. (**1986**). "Psychophysical suppression of selective portions of pulsation threshold patterns," Hear. Res. **21**, 257–260.

Wegel, R. L., and Lane, C. E. (**1924**). "The auditory masking of one pure tone by another and its possible relation to the dynamics of the inner ear," Phys. Rev. **23**, 266–285.

Wojtczak, M., and Viemeister, N. F. (**2005**). "Psychophysical response growth under suppression," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer, New York), pp. 67–74.

Yasin, I., and Plack, C. J. (**2003**). "The effects of a high-frequency suppressor on tuning curves and derived basilar-membrane response function," J. Acoust. Soc. Am. **114**, 322–332.

Yasin, I., and Plack, C. J. (**2005**). "The role of suppression in the upward spread of masking," J. Assoc. Res. Otolaryngol. **6**, 368–377.

Yates, G. K. (**1990**). "Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions," Hear. Res. **50**, 145–162.

Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). "BM nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hear. Res. **45**, 203–220.

Zinn, C., Maier, H., Zenner, H-P., and Gummer, A. W. (**2000**). "Evidence for active, nonlinear, negative feedback in the vibration response of the apical region of the in-vivo guinea-pig cochlea," Hear. Res. **142**, 159–183.

Auditory short-term memory persistence for tonal signals in a songbird

Melanie A. Zokoll, Georg M. Klump, and Ulrike Langemann[a]
Zoophysiology and Behavior Group, Institute for Biology and Environmental Sciences, Carl von Ossietzky University, Postfach 2503, 26111 Oldenburg, Germany

This paper presents an animal model for studying the persistence of auditory memory for tonal signals. Five European starlings (*Sturnus vulgaris*) were trained in a Go/NoGo delayed nonmatching-to-sample task to discriminate between a series of identical "sample stimuli" and a single "test stimulus." Frequencies of sample and test stimuli should be classified as being either the same or different. The performance of the birds is measured as the percentage of correct classifications. Three parameters were varied: The delay between sample and test stimuli, the number of sample stimuli presented before the test stimulus, and the salience of the difference between sample and test stimuli. Auditory memory persistence time was estimated as a function of the delay between the last sample stimulus and the test stimulus. The performance of the birds deteriorated with increasing delay before the test stimulus. Increasing the number of sample stimuli in the series of tones presented before the test stimulus improved performance. Performance was also better for more salient differences between sample and test stimuli. The individual auditory memory persistence time varied between 4 and 20 s.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2713721]

I. INTRODUCTION

To evaluate a sequence of acoustic communication signals it is necessary to keep previous elements in memory. In humans, short-term storage of sounds allows, for example, the discrimination of the pitch of successive signals or the maintenance of sentence comprehension by analyzing successive speech elements (e.g., Schröger and Berti, 2000). Concepts of memory performance are often explicitly language related (e.g., "phonological loop", Baddeley, 2003). The need for storing information not only applies to humans, but should be common for all species using sequences of sound signals. Thus, animal models can help to clarify the universal validity of memory concepts.

The structure of vocal signals in songbirds has striking parallels to human speech organization and both song and speech are acquired by vocal learning (e.g., Doupe and Kuhl, 1999). In the European starling (*Sturnus vulgaris*), for example, both males and females are adept singers with complex, multimotif songs that last for up to 50 s (Eens *et al.*, 1991). Female starlings base their mate choice on the diversity of elements that an individual male sings (e.g., Eens *et al.*, 1991; Mountjoy and Lemon, 1996). To assess the diversity of ongoing songs, a female needs to store information about the song elements for a short time period. The same applies to a male starling judging the diversity of a rival's song. Thus, songbirds with a large repertoire of vocal signals like the starling require memory mechanisms to analyze their songs similar to humans when analyzing speech.

In animal studies both delayed matching-to-sample (DMTS) and delayed nonmatching-to-sample (DNMTS) are established procedures that are commonly used to assess memory persistence time (e.g., Pasternak and Greenlee, 2005; Paule *et al.*, 1998). In a typical DMTS experiment, trials consist of the presentation of a sample stimulus which is separated from one or two comparison test stimuli by a delay. The test stimulus that matches the sample stimulus is the correct (rewarded) choice. In the DNMTS variant, the nonmatching stimulus is the correct choice. At any single trial subjects cannot determine the correct response before the presentation of the test stimulus, i.e., some signal characteristics of the sample stimulus (e.g., frequency) must be memorized by the subject in order to respond correctly (Gottlieb *et al.*, 1988; Konorski, 1959). To assess memory persistence, the delay between sample and test is increased, whereupon choice accuracy at the stage of the comparison usually diminishes (e.g., Cowan, 1984; White, 2001; White and Wixted, 1999). This decline in choice accuracy with increasing delay indicates that subjects "forget" the information obtained from the sample stimulus (Sargisson and White, 2003). The auditory memory persistence time can be estimated as the delay between the sample and the test stimulus at which the performance drops to a certain threshold value (e.g., 75% correct responses).

In the present study, a Go/NoGo DNMTS paradigm was applied to estimate the persistence time of the starlings' auditory memory. Since we wanted to investigate basic memory phenomena we decided to probe the starlings' auditory memory persistence time with artificial pure tone stimuli that differed in frequency. Using species specific song might have caused special processing [e.g., in the "song system" of birds, see Brainard and Doupe (2002), or by vocalization-

---

[a]Author to whom correspondence should be addressed; electronic mail: ulrike.langemann@uni-oldenburg.de

selective neurons in auditory forebrain regions, e.g., Gentner *et al.* (2004), Gentner and Margoliash (2003)]. Frequency is a particularly important feature of many bird communication signals (e.g., Dooling, 2004). Since starlings usually repeat motifs in a song several times before singing the next motif (e.g., Adret-Hausberger *et al.*, 1990), this repetition could have a behavioral function to produce a more stable memory for single motifs helping the receiver to judge song diversity. One of our goals was to test whether the repetition of tone samples prior to the test stimulus presentation affects a bird's memory performance. The salience of the nonmatching cue, i.e., the size of the difference between sample and test stimuli, is another factor that should influence memory performance (e.g., Bachem, 1954; Schröger, 1997) and that was investigated in the present experiment.

II. MATERIALS AND METHODS

A. Subjects

Five wild-caught European starlings (*Sturnus vulgaris*, three females: Gi, Li, Ma, and two males: An, Ru) between 6 and 10 years old served as the subjects. Two of the starlings had previously participated in psychoacoustic discrimination experiments. The birds were housed in individual cages of $80 \times 40 \times 40$ cm^3 with at least 14 light hours. The starlings were maintained at 86%–97% of their free-feeding weight on a diet of duck-food pellets [Mastvollkorn Geflügel Finisher, Raiffeisen Central-Genossenschaft Nordwest eG (RCG), Germany]. Water was available at all times in the home cages. Food rewards during the experiments consisted of pieces of mealworms (larvae of *Tenebrio molitor*), which were among the birds' favorite foods.

The care and treatment of the birds were in accordance with the procedures of animal experimentation approved by the Government of Lower Saxony, Germany. All procedures were performed in compliance with the NIH Guide on Methods and Welfare Considerations in Behavioural Research with Animals (NIH Publication No. 02-5083).

B. Apparatus

The birds were tested in an experimental cage ($24 \times 36 \times 32$ cm^3) constructed from wire mesh which was placed in a sound-attenuated echo-reduced box. The box was lined with sound-absorbing wedges (Illbruck Illsonic Pyramide 100/100 mounted on 50 mm of Illsonic Plano, cutoff frequency of 500 Hz, $\alpha > 0.99$; Illbruck GmbH, Germany). On the front wall of the experimental cage two response keys (observation key and report key) each with a light-emitting diode (LED) and a light-interrupting switch were attached. Below the keys a rotary food dispenser operated by a stepping motor was mounted, which delivered the food rewards to the bird during the experiment.

C. Stimulus generation

Pure-tone signals were generated using a Linux-operated workstation (AMD Processor, Sound Blaster PCI 512, 44.1 kHz sampling rate). The signals were adjusted in level by a computer-controlled attenuator (TDT PA4; Tucker-
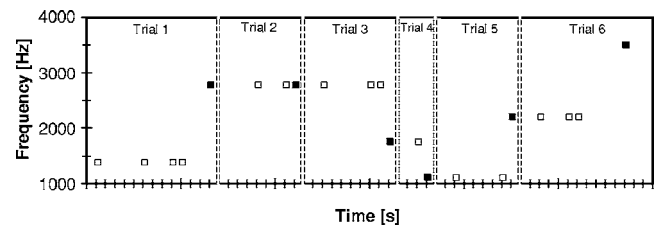
FIG. 1. Exemplary sequence of six trials. Units at the *x* axis represent seconds elapsed within each trial. The frequencies of the sample tones (open squares) and test tones (closed squares) either were the same (trial 2) or they were different (trial 1, and trials 3–6). Each trial contained up to six sample stimuli and a final test stimulus presented with one delay chosen at random from the three different delays of a delay set. The delays between sample stimuli were chosen randomly from the same delay set.

Davies Technologies, Alachua, FL) before being amplified (Yamaha AX-500) and played by the loudspeaker (Canton Twin 700, 200-9000 Hz, ±2.5 dB; Canton Elektronik, Germany) in the sound-proof chamber. The speaker was positioned approximately 30 cm above and slightly behind (6 cm) the bird's head. Sound pressure levels were measured (General Radio 1982 Precision Sound Level Meter & Analyzer with the 1/2 In.-condenser microphone) at approximately the positions where the bird's head would be during the experiment. The level of a stimulus of a given frequency differed by less than 2.5 dB between possible positions of the bird's head, and the level at different frequencies varied by less than 3.4 dB at any given position.

All auditory stimuli were pure tones with frequencies between 1 and 4 kHz (i.e., 1100, 1386, 1746, 2200, 2772, and 3492 Hz), which is well within the frequency limits of perception in the starling (see Langemann *et al.*, 1999). The stimuli had a minimum frequency difference of a third of an octave, i.e., the frequency differences were at least about 20 times larger than the starling's frequency difference limen [FDL, e.g., Langemann and Klump (1992)] and should have been salient to the birds. The total duration of the stimuli was 400 ms, including 8 ms Hanning ramps at stimulus onset and offset. Pure tones were presented at a mean overall level of 61 dB SPL, which corresponds to a sensation level of about 51–56 dB in the starling for frequencies of between 1 and 4 kHz (Langemann *et al.*, 1999). To prevent the birds from possibly using perceived loudness changes instead of frequency changes (e.g., Moore and Glasberg, 1989), the amplitude of the stimuli was randomized from one stimulus presentation to the next in the range of ±3 dB.

D. Procedures

The starlings were trained in a Go/NoGo variant of the DNMTS procedure. The behavioral protocol was controlled by a Linux-operated workstation. Pecking an observation key at least two times initiated the presentation of up to six sample stimuli (see open squares in Fig. 1) of equal frequency with random interstimulus intervals of between 1 and 24 s (see delay sets in Table I). In the example given in Fig. 1, four sample stimuli with different delays of 5, 3, and 1 s were presented in the first trial, two sample stimuli in the second trial, and three sample stimuli in the third trial, etc. (delays were chosen from the specific set, e.g., 1, 3, and 5 s).

TABLE I. The delay sets with which subjects were tested, the number of subjects tested (N_{subjects}), and the maximum number of sample stimuli presented in trials of that delay set (N_{sample}). With increasing delay set number fewer sample stimuli were presented before the test stimulus to prevent single trials from becoming too long. When performance of a starling was still above 75% correct for the longest delay of a set, the next delay set was tested. Thus, as the delay was increased, fewer starlings were tested.

Delay set	N_{subjects}	Delay (s)	N_{sample}
1	5	1, 2, 3	6
2	5	1, 3, 5	6
3	5	2, 4, 6	6
4	3	3, 6, 9	4
5	2	4, 8, 12	3
6	2	6, 12, 18	3
7	1	8, 16, 24	3

After a further delay in the same temporal range as the delays between sample stimuli, a test stimulus was presented, indicated by closed squares in Fig. 1. A test stimulus had either the same frequency (matching or NoGo stimulus) or a different frequency (nonmatching or Go stimulus) as the sample stimuli. The frequency of the test stimulus of the previous trial was used as the repeating sample stimulus of the new trial. Half of the trials consisted of sample test stimulus combinations with matching frequencies (correct behavior was withholding a response, i.e., correct rejection) and the other half of the trials included sample test stimulus combinations that were nonmatching in frequency. If the bird pecked the report key within 3 s from the presentation of the test stimulus with the nonmatching frequency, the LEDs behind the pecking keys were extinguished, the food tray rotated, and the bird was reinforced with a piece of mealworm for making a hit. A feeder light was presented as a secondary reinforcer. A response to the report key during the presentation of a matching stimulus (i.e., a false alarm) was followed by a timeout period of between 3 and 20 s with the LEDs behind the pecking keys and the lights in the experimental cage switched off. The duration of the timeout period was adjusted in order to control the false alarm rate of the birds. To prevent changing the delays between sample stimuli, pecking the report key during the presentation of the sample stimuli produced no timeout. Withholding a response to the presentation of a test stimulus with nonmatching frequency (i.e., a miss) had no effect. In case of a miss or a hit the next trial started after an intertrial interval of 3 s and the bird could initiate the presentation of the next stimulus by pecking the observation key at least two times.

Since the experimental runs progressed over several weeks with a stepwise increase in the delays tested, we felt the need to check for possible training effects on performance. To assess if performance would change over time, we presented the birds with a new frequency set in a subsequent experiment and determined auditory memory persistence time again. In this experiment the frequencies were 1235, 1556, 1960, 2470, 3112, and 3921 Hz. The experiment was run with the last delay set presented in the previous experiment, i.e., that delay combination at which the bird's performance for at least the longest delay dropped below the threshold criterion of 75% (see the following).

E. Estimate of memory performance and data analysis

Memory performance was estimated by providing the subjects with trials organized in matrices of six different stimulus frequencies that were presented with three different delays (see delay sets in Table I, e.g., sets contained delays of 1, 2, and 3 s, or 3, 6, and 9 s, etc.). With 6 different frequencies 30 nonmatching stimulus combinations are possible. Since each combination was presented with three different delays, this resulted in 90 nonmatching sample test stimulus combinations per matrix. An additional 90 trials in the same matrix included sample test stimulus combinations with matching frequencies at a similar distribution of delays. A total of 180 trials per matrix was thus presented in a pseudorandom order constrained by the rule that any test stimulus will become the sample stimulus of the following trial. All starlings had to complete three matrices of 180 trials each (i.e., 540 trials in total) with a given delay set. The combinations of the three different delays presented in a matrix were chosen to provide a minimum delay at which the subject would perform above threshold and two longer delays. Delays between the sample stimuli as well as between the sample and test stimuli were randomly chosen from the delay set. On average all delays within a specific delay set were presented with equal probability.

After completion of 540 trials with a given delay set, the starlings were presented with the next more difficult delay set (see Table I), until their performance for at least the longest delay in the set dropped below the threshold criterion. Birds that reached this point were not tested further with longer delays. Memory performance was expressed as the ratio between correct responses and all possible responses:

$$\text{correct responses } [\%] = 100\left(\frac{c_n + c_m}{c_n + i_n + c_m + i_m}\right), \qquad (1)$$

where c_n and c_m are the number of correct responses in nonmatching (i.e., hits) and matching trials (i.e., correct rejections), respectively. The number of incorrect responses in nonmatching (i.e., misses) and matching trials (i.e., false alarms) are i_n and i_m, respectively. Auditory memory persistence time was calculated by linearly interpolating the two delays at which the starlings performed just above and below the threshold criterion of 75% correct responses, respectively. Data were combined from more than one delay set for threshold estimation. To control the starling's motivation and attention, we discarded matrices in which a bird's memory performance for the shortest delay did not reach at least 75% correct responses.

For evaluating the birds' performance in relation to the salience of the frequency difference, the hit rate in nonmatching trials was calculated according to

$$\text{hit rate } [\%] = 100\left(\frac{c_n}{c_n + i_n}\right). \qquad (2)$$
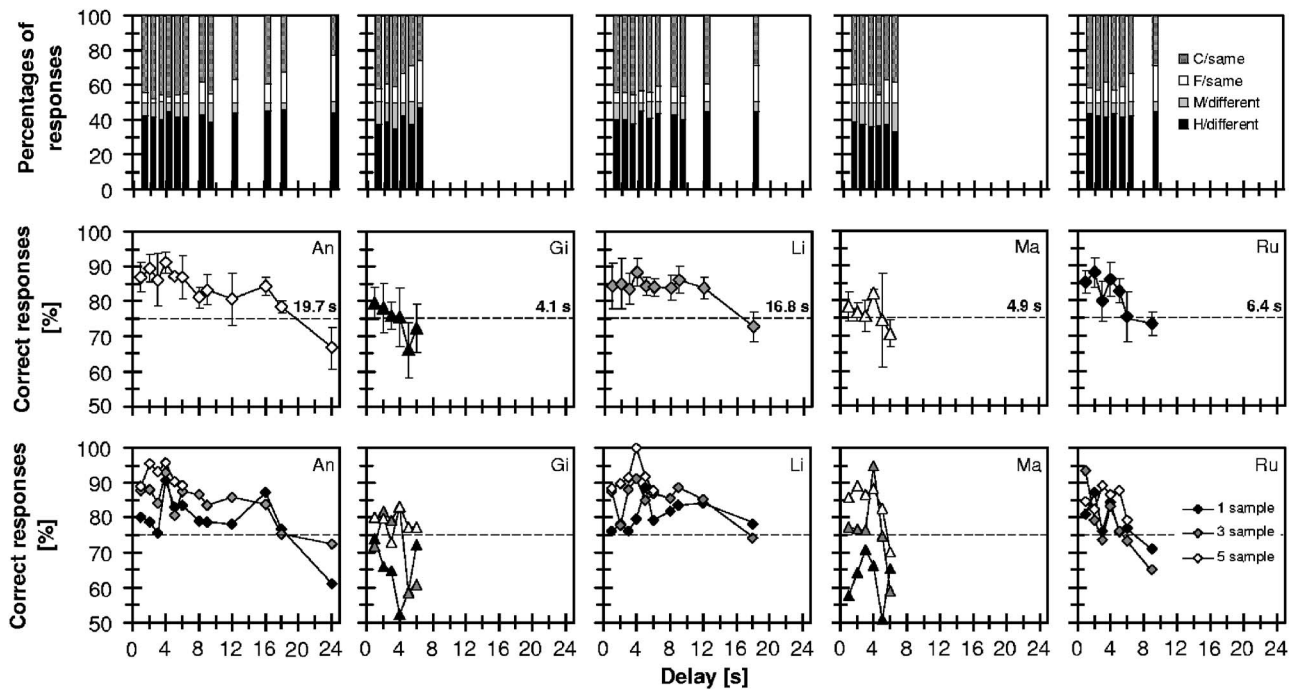
FIG. 2. Individual performance of five birds as a function of the delay between the last sample and the test stimulus. **Top row:** Mean percentages of all responses. Hit (H) or miss (M) can be scored if sample and test stimuli differ in frequency; correct rejection (C) or false alarm (F) can be scored if sample and test stimuli frequency are matching. **Middle row:** Mean performance ($\pm$ s.d.) for each of the five starling individuals. Individual performances are expressed as percent correct responses, i.e., $100((c_n+c_m)/(c_n+i_n+c_m+i_m))$, where c_n and c_m are the number of correct responses in nonmatching (i.e., hits) and matching trials (i.e., correct rejections), respectively. The number of incorrect responses in nonmatching (i.e., misses) and matching trials (i.e., false alarms) are i_n and i_m, respectively. The dotted line indicates the threshold criterion of 75% correct responses and the average auditory memory persistence time of the individual starling (i.e., the delay at which the performance dropped to 75%) is given in each panel. **Bottom row:** Mean performance for the five starlings for three of the six different sample stimulus repetitions (1, 3, and 5 sample stimuli).

III. RESULTS

To evaluate the effects of the parameters frequency difference (in third octave steps), delay (delays for which all five starlings were tested), and number of sample repetitions on the birds' auditory memory persistence, we have analyzed the hit rates of the birds in nonmatching trials with an ANOVA (general linear mixed model). Frequency difference, delay, and the number of sample repetitions were incorporated in the analysis as fixed factors, and the identity of the individual birds was included as a random factor. The dependent variable was the hit rate. Viewed across all conditions there were significant main effects of all three fixed factors (frequency difference $F=15.477$, $P<0.001$; delay $F=4.761$, $P<0.001$; number of repetitions $F=11.900$, $P<0.001$). The ANOVA indicated no difference between the individuals' responses. There were no significant interactions between any of these factors. In the following three sections we describe the effect of each single factor separately and include a detailed description of overall performance in both matching and nonmatching trials.

A. Effect of delay

The starling's auditory memory persistence time was determined on the basis of the overall performance (i.e., percentage of correct Go or No-Go-responses indicating whether sample and test stimuli differed or were the same, respectively). Data for a given delay obtained in the different delay sets were averaged for each bird. The percentage of

correct responses is seen as an indicator for the strength of the memory trace. The middle row of Fig. 2 shows the mean memory performance of each of the five starlings as a function of the delays tested. Data points are based on the analysis of 180 trials in the case that the delay was only contained in one delay set, or they are based on up to 540 trials if the delay occurred in three different delay sets (Table I). Memory performance was calculated separately for each matrix and the data were averaged across matrices. The mean memory performance of each of the individuals was well above the chance level of 50% correct responses for all delays tested with the exception of the longest delays. At short delays of 1 to 2 s, three of the birds (An, Li, Ru) showed a mean memory performance that was above 80% correct responses. The other two birds showed a mean memory performance of between 77% and 79% correct responses, which was already quite close to the threshold criterion of 75% correct responses. In general, the percentage of correct responses decreased with increasing delay for all birds (see Fig. 2, middle row).

To test if the starlings' performance in matching and nonmatching trials varied significantly with delay, we performed a one-way repeated measures analysis of variance (one-way RM ANOVA) with the individual mean percentage of correct responses as the dependent variable. Since the birds were tested with a different range of delays, we included in the ANOVA only the mean memory performance for the six longest delays for which any individual bird was tested and ranked the delays by number one to six (indepen-
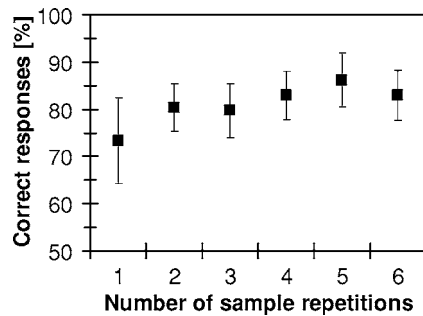
FIG. 3. Mean performance of five starlings [percent correct responses, i.e., $100((c_n+c_m)/(c_n+i_n+c_m+i_m))$] as a function of the number of repetitions of the sample stimulus (1–6). The number of correct responses in nonmatching (i.e., hits) and matching trials (i.e., correct rejections) are c_n and c_m, respectively. The number of incorrect responses in nonmatching (i.e., misses) and matching trials (i.e., false alarms) are i_n and i_m, respectively. Only data from delay sets for which performance for all five starlings was determined are shown.

dent variable). The mean memory performance for the six longest delays for each individual bird revealed statistically significant differences ($F=9.816$; $P<0.001$). Tukey post-hoc tests showed that the starlings performed significantly worse for the longest delay tested than for every other delay except for the second longest delay (all $P<0.050$).

The individual starlings had very different auditory memory persistence times: While three of the five starlings tested showed auditory memory persistence times of between 4 and 7 s (Fig. 2, middle row), two of the starlings had auditory memory persistence times above 16 s (An, Li). The individual starlings' auditory memory persistence times calculated by linear interpolation of the mean memory performance for the different delays were 19.7 s (An), 4.1 s (Gi), 16.8 s (Li), 4.9 s (Ma), and 6.4 s (Ru; Fig. 2, middle row). The top row in Fig. 2 shows the percentages of the different responses in nonmatching and matching trials (hit, miss, correct rejection, and false alarm) for the various delays tested. The percentages of responses in the threshold range of the individual birds indicate that the mean memory performance dropped mainly because of the birds' increasing false alarms. An additional one-way RM ANOVA with percentage of false alarms as the factor was performed to test if the starlings' mean percentage of false alarms differed significantly for the six longest delays. This test revealed that the percentage of false alarms was significantly different for the six longest delays of each individual starling ($F=5.015$; $P=0.004$). The percentage of false alarms was highest at the longest delay, and was significantly larger than all other delays (Tukey tests; all $P<0.050$) apart from the delay that was just shorter than the longest delay.

B. Effect of sample stimulus repetition

In the experiments, the sample stimulus was presented repeatedly before the test stimulus. Thus, the effect of sample stimulus repetition on mean memory performance could be investigated. Figure 3 shows the five birds' mean memory performance as a function of the different numbers of repetitions of the sample stimulus presented before the test stimulus. This comparison was based on data obtained in the first

three delay sets (1620 trials for each bird), for which all five starlings were tested (cf. Table I) and that included presenting up to six sample stimuli. A one-way repeated measures ANOVA with number of sample stimulus repetitions (1–6) as the factor was performed. The five starlings' mean memory performance for the different sample stimulus repetitions was significantly different ($F=10.301$; $P<0.001$). The worst mean memory performance was observed for trials in which only one sample stimulus was presented before the test stimulus with a mean correct response of 73.4% for the five birds (Tukey post-hoc tests, $P=0.018$ for 1 vs 2 sample stimuli; $P=0.040$ for 1 vs 3 sample stimuli; and all $P \leq 0.001$ for 1 vs 4, 1 vs 5, and 1 vs 6 sample stimuli). The mean memory performance for the different numbers of repetitions greater than one was not significantly different (with one exception of a significant difference between the mean memory performance for three sample stimuli in comparison to five sample stimuli presented before the test stimulus, Tukey post-hoc test, $P=0.028$). Thus, the starlings performed better with any repetition of the sample stimulus before the test stimulus compared to no repetition. To demonstrate the change in the individual memory performance with sample stimulus repetition, mean memory performance with a different number of sample stimuli presented before the test stimulus is depicted in Fig. 2 (bottom row) as a function of delay. For a better overview, individual memory performance is shown only for conditions with 1, 3, and 5 sample stimuli. The mean memory performance for 3 and 5 sample stimuli presented before the test stimulus was most of the time superior to the condition with only one sample stimulus presented before the test stimulus. If only one sample stimulus was presented before the test stimulus, the mean performance of two of the five starlings (Gi, Ma) was below the threshold criterion of 75% correct responses even for the short delays. The graphs in the bottom row of Fig. 2 show that the results of the two birds depicted in the middle row of Fig. 2 crucially depended on the repetition of the sample stimulus prior to test stimulus presentation. Only by means of repeating the sample stimulus were they able to perform above the threshold criterion which was necessary to calculate an auditory memory persistence time. The remaining three starlings still performed above threshold for delay values of up to 7.0 (Ru), 18.7 (An), and 21.3 s (Li, estimated by extrapolating the data) with only one sample stimulus presented before the test stimulus.

C. Effect of frequency difference between sample and test stimulus

To test whether the salience of the frequency differences between sample and test stimuli influences performance, the starlings were presented with six tone stimuli differing in frequency by a minimum difference of one-third octave and a maximum difference of five-thirds of an octave. Our general analysis using an ANOVA that is presented at the beginning of the results already has shown that the frequency difference measured in third-octave steps significantly affects the hit rate. Since the starling's cochlear excitation patterns for the different frequencies at the signal level used in this study broadly overlap (see Buus et al., 1995), we grouped
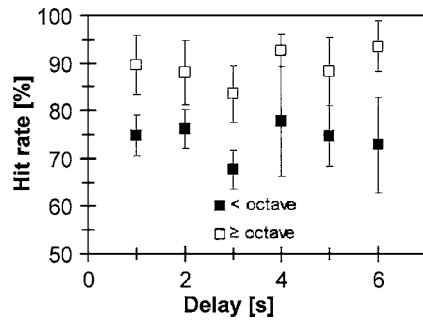
FIG. 4. Mean performance (± s.d.) of the five starlings [hit rate [%] $=100(c_n/(c_n+i_n))$] for frequency separations smaller than one octave and equal to or larger than one octave as a function of the different delays. The number of correct (i.e., hits) and incorrect responses (i.e., misses) in non-matching trials are c_n and i_n, respectively. Only delay conditions in which hit rates for all five starlings were acquired are shown. Differences between the two frequency difference categories were significant for all different delays.

our data of the mean hit rates [calculated according to Eq. (2)] in the further analysis into two classes: Frequency differences smaller than one octave, and frequency differences equal to or larger than one octave. Data from the first three delay sets (cf. Table I) that included measurements on all five starlings were combined. Figure 4 shows the mean hit rate as a function of the size of the frequency difference. We tested whether the mean performances for the two frequency-difference classes were significantly different for the various delays that all five starlings could be tested with (1–6 s). The mean performance was significantly worse (paired t-tests, all $P < 0.05$) for frequency differences smaller than one octave than for frequency differences of at least one octave. Averaged across the different delays, the starlings' mean performance for frequency differences smaller than one octave was 74.0% (±4.9%, s.d.), whereas for frequency differences larger or equal to one octave their mean performance was 89.2% (±3.8%).

D. Effects of training on auditory memory persistence times

To test if the birds' performance in the experiment was influenced by presenting the same set of frequencies repeatedly with different delays (Table I), we compared the auditory memory persistence times obtained in this experiment with those obtained in a subsequent experiment using a second set of frequencies that was new to the birds. At the start of this experiment the delays were chosen to be similar to those used in the last matrix of trials that determined memory performance with the first set of frequencies. Table II compares the individual auditory memory persistence

TABLE II. Individual auditory memory persistence times in seconds of five starlings (An, Gi, Li, Ma, and Ru) for the two different frequency sets (frequency set 1: 1100, 1386, 1746, 2200, 2772, 3492 Hz; frequency set 2: 1235, 1556, 1960, 2470, 3112, 3921 Hz) presented in the first and second experiments.

Frequency set	An	Gi	Li	Ma	Ru
1	19.7	4.1	16.8	4.9	6.4
2	19.6	4.6	12.7	11.7	6.6

times of the birds. There was no significant difference between auditory memory persistence times acquired in the two experiments (paired t-test, $P > 0.05$). In three starlings, auditory memory persistence times for the first and second experiments were approximately the same (An, Gi, and Ru). Two birds showed a larger change in auditory memory persistence times, but the changes were in opposite directions between the two experimental series.

IV. DISCUSSION

A. Effect of delay

Auditory memory persistence times for the individual starlings span a broad time range from 4.1 to 19.7 s (see Fig. 2, middle row). While three of the five starlings showed relatively short auditory memory persistence times of 4.1, 4.9, and 6.4 s, two of the starlings had quite long auditory memory persistence times of 16.8 and 19.7 s. For all birds in this study, the mean percent correct responses declined as the delay was increased from 1 s to the maximum delay that could be tested in each bird (6 to 24 s, Fig. 2, middle row). The decline in performance over time which has previously been termed "forgetting curve" or "retention gradient" (Paule et al., 1998) is thought to indicate the rate of decay of the sensory trace of the sample stimulus.

The change of auditory memory over time has also been studied in other animals applying DMTS/DNMTS paradigms. The only other bird species in which sensory memory has been studied in an auditory DNMTS task is the budgerigar (*Melopsittacus undulatus*, Downing et al., 1988). Compared to the budgerigars in the study of Downing et al. (1988), starlings show a rather long auditory memory persistence time. While the budgerigar's performance reached chance level at a delay of 2 s (Downing et al., 1988), starlings still maintained a mean performance of 83.5% (±5.8%) correct responses at a delay of 2 s. The study on budgerigars was conducted with only one sample stimulus presented before the test stimulus, as is often done in DMTS/DNMTS experiments (e.g., Paule et al., 1998), whereas in this study up to six sample stimuli were presented before the test stimulus.

For a better comparison of the performance of starlings with that of other species, performance for the different delays in the starling was calculated for conditions with only one sample stimulus presented before the test stimulus. Performance and memory persistence times in the other species were estimated from the published data applying the same threshold criterion as in the present study. Figure 5 shows the data of auditory DMTS/DNMTS experiments on different animal species compared to the data from the present experiment on starlings. The starlings were divided into two groups: Subjects with long auditory memory persistence times (An, Li), and subjects with short auditory memory persistence times (Gi, Ma, and Ru). Data are only shown for delays for which results from at least two birds were available.

The budgerigar's auditory memory persistence was 0.54 s averaged for both budgerigars tested (range 0.4–0.8 s) and including data from all three sets of stimuli [a
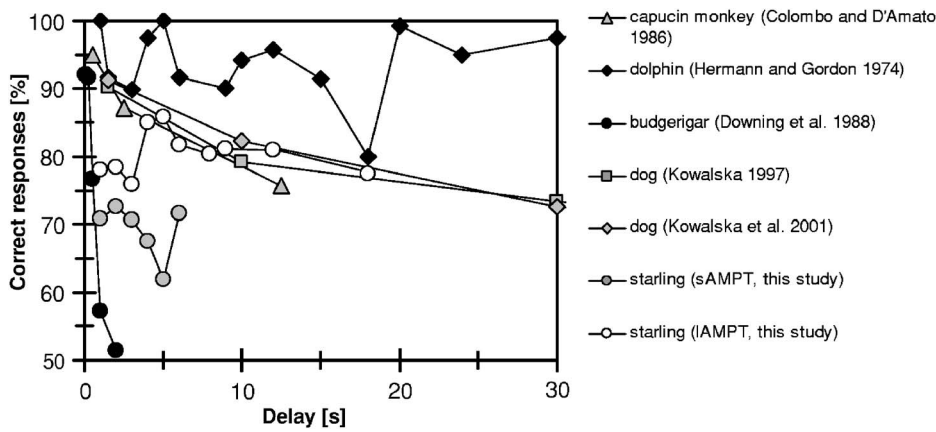
FIG. 5. Mean performance for different animal species expressed as percent correct responses for each delay. Starlings with long auditory memory persistence time (AMPT) were grouped into the lAMPT group, and starlings that showed a comparatively short AMPT were grouped into the sAMPT group. Since only one sample stimulus was presented before the test stimulus in the compiled studies, only data from trials in which the starlings were presented with one sample stimulus were included.

pair of tonal signals, a pair of budgerigar calls, and a tone-call-pair, Downing *et al.* (1988)]. This is still considerably shorter than the long auditory memory persistence times shown in two of the starlings (i.e., memory persistence times differ by a factor of up to 30). Comparable studies in mammals (using various acoustic stimuli) demonstrated long auditory memory persistence and large interindividual differences in auditory memory persistence times similar to the starling. Capucin monkeys (*Cebus apella*) showed auditory memory persistence times ranging from 3 to 12.5 s for tonal and complex signals [Colombo and D'Amato (1986), four subjects]. In comparison, dogs (*Canis lupus familiaris*) show longer but also individually quite different auditory memory persistence times, with performance dropping below 75% correct responses at delays ranging from 7.6 to 96 s [320 different sounds served as stimuli; Kowalska (1997), three subjects; Kowalska *et al.* (2001), twelve subjects]. In an auditory memory experiment by Herman and Gordon (1974), the dolphin (*Tursiops truncatus*) tested showed the longest auditory memory persistence time, since for delays up to 120 s performance showed no decay and never dropped below a correct response value of 75%.

In humans, estimates of the duration of the "long" auditory memory store have varied extensively depending on the experimental paradigm (for reviews see Cowan 1984, 1997). Tones with frequency differences near the FDL (Wier *et al.*, 1977) were found to be stored in auditory memory for at least 10 s (Keller *et al.*, 1995; Cowan *et al.* 1997; Clément *et al.*, 1999; Koester, 1945) which is in the same range as the memory persistence time observed in the starling. Longer memory persistence times of up to one week were observed with larger frequency differences (e.g., >10%) and the ability to perceive absolute pitch affected performance over long retention times (Bachem, 1954). Two additional studies in which human subjects were distracted (by reading or by intervening tones) revealed auditory memory persistence times for tones of between 7 and 12 s (Sams *et al.*, 1993; Wickelgren, 1969). There is also neuromagnetic evidence supporting a neuronal representation of tones for at least 10 s in humans (Böttcher-Gandor and Ullsperger, 1992; Sams *et al.*, 1993). Sams *et al.* (1993) proposed that the delay dependence of the behavioral performance and that of the neuromagnetic response investigated in their study are similar enough to suggest a common underlying mechanism. In

summary, the starling's auditory memory persistence time lies within the range of the results observed in other vertebrate species.

Durlach and Braida (1969) suggested a model of auditory discrimination of successive stimuli that could explain the decay of performance with increasing delay between the two stimuli. They suggested that the stimulus is transformed to a sensory trace that is degraded by continuously adding noise to the sensory trace over time (also the initial transformation is noisy, but this is not important in the discussion of the effect of delays). In the discrimination, the degraded sensory trace of the first stimulus is compared to the sensory trace of the more recent second stimulus by computing the difference between the two representations. Kidd *et al.* (1988) suggested that this simple model can be applied to any discrimination of spectra (e.g., signals involving multiple channels). The model by Durlach and Braida (1969) can also be used to explain the results in the present DNMTS study. The sample tone generates a sensory trace that degrades over time, and this degraded representation is compared to the representation of the test tone. The larger the delay the more degradation has occurred and the more is the sensory trace affected by the noise that is added during the delay rather than the sample stimulus. After a long delay the difference between the two representations does not depend on the sample stimulus representation any more and, therefore no differential response to the various test tone frequencies will occur that depends on the previous sample stimulus frequency. Thus, only chance performance is possible. If only a short delay is introduced between sample and test stimulus, the various test tone frequencies will provide for a differential response that depends on the previous sample tone representation. In this case the difference between the two representations is varied depending on the relation between sample and test tone frequency and a good performance is possible.

Previous animal studies suggest that auditory memory persistence may not be purely due to a sensory effect. Weavers *et al.* (1998) wondered if the decrease in matching accuracy over time could be at least partly accounted for by an effect of delayed reinforcement on sample-directed remembering behavior rather than the sensory representation of the stimulus. There are experiments in which these two aspects were separated to some extent (e.g., McCarthy and Davison,

1991; Weavers *et al.*, 1998) and the results suggested that the decline in performance may be partly accounted for by lack of immediate reinforcement for remembering. In the present experiment, as in most DMTS/DNMTS procedures, the delay to choice and the delay to reinforcement are essentially the same (e.g., Downing *et al.*, 1988; Kowalska, 1997; Wright *et al.*, 1990) and, therefore, the effect that delayed reinforcement could have had on the results cannot be evaluated. In general, the starlings show an increasing false alarm rate with longer delays (Fig. 2 top row) indicating an increasing proclivity to make a behavioral response with longer time intervals that is unrelated to the stimuli, but that might be associated to the expected reinforcement (e.g., Staddon, 1983).

B. Effect of sample stimulus repetition

Our experimental design was suitable to additionally test whether the stability of the auditory memory trace formed by the sample stimuli is dependent on their recurrence. Figure 3 shows the starling's average performance as a function of the different repetitions of sample stimuli presented before the test stimulus. Since the birds perform significantly better if more than one sample stimulus is presented before the test stimulus, the results suggest that repetition of a given stimulus leads to a more stable auditory memory trace. It seems that a single repetition of the sample stimulus is sufficient to reach the maximum level of performance, since further repetition in general does not lead to a statistically significant improvement in performance (i.e., there is no strong cumulative effect beyond one repetition). These findings highlight the importance of at least one repetition for the study species: The birds with short auditory memory persistence times only fulfilled the criterion of 75% correct responses if conditions with repetition were included in threshold determination in which more than one sample stimulus was presented before the test stimulus (see Fig. 2 bottom row).

The results of the present study may be compared to results from two studies of Deutsch (1970, 1972a) in humans. She asked subjects to judge whether a sample and a test tone stimulus were identical or different, but presented intervening tones varying in frequency between sample and test stimulus. If one of the intervening tones was identical in frequency to the sample stimulus the subject's performance was improved (i.e., they showed a decreased error rate) indicating the benefit gained from repetition. Deutsch (1970) proposed that this finding could have resulted from a trace strengthening effect of repeating the sample tone. Other studies on human subjects investigated the influence of "controlled rehearsal" (i.e., an active process) on memory performance and decay. In a study by Peterson and Peterson (1959) using verbal items forgetting progressed at different rates depending on the amount of controlled rehearsal. Peterson and Peterson (1959) provided evidence that rehearsal does not merely postpone the onset of forgetting, but the forgetting curve had an asymptote dependent on the number of rehearsals (see also Peterson, 1982). Improvement with rehearsal also seems to work for tonal stimuli in humans. Keller *et al.* (1995) reported better performance in a two-tone

comparison task in which participants were instructed to rehearse the pitch of the first tone covertly during the delay, compared to conditions in which such rehearsal was prevented by a distractor task. The rehearsal process is thereby seen as "auditory imagery" that can be used strategically to reduce the rate of decay of auditory information for tone pitch (Keller *et al.*, 1995). Auditory imagery in turn seems to share common neural substrates with actual listening (Yoo *et al.*, 2001).

In one of her studies, Deutsch (1972b) tried to relate the signals that were presented to the subjects to their percept. Deutsch proposed a hypothesis explaining the observed memory for pitch of a tone and its decay in time. Her model resembles the model proposed by Durlach and Braida (1969) introduced above. The stimuli are represented in form of a three-dimensional bell-shaped central distribution of excitation reflecting the memory trace in which the *z*- and *x*-axes represent frequency and ongoing time, respectively, and the *y*-axis the strength of excitation. As time proceeds, the strength of the excitation decreases. Repetition of the sample stimulus increases the strength of its representation and thus provides for a larger difference in the representation between the sample stimulus and the test stimulus. Additional stimulation with the same frequency thus leads to summed activity taking longer to decay to a threshold level than activity resulting from just one stimulation. This indicates that consolidation of memory through repetition results in better discrimination performance (Deutsch, 1972b, 1975).

Stimulus repetition also may play an important role in the context of starling natural behavior. Male starlings tend to sing in long continuous episodes, termed bouts. Song bouts, in turn, are composed of much smaller acoustic units referred to as motifs (Adret-Hausberger and Jenkins, 1988; Eens *et al.*, 1991). The motifs are usually highly stereotyped between successive renditions. The starlings' behavior of repeating a certain motif in a song several times before singing the next motif (e.g., Adret-Hausberger *et al.*, 1990) could be an adaptation to allow female starlings a better evaluation of the male's complicated, multimotif songs which last up to 50 s. The present study demonstrates that repetition will increase auditory memory performance. The repetition might thus help the females in judging song quality by strengthening the memory traces for the different song motifs that have to be compared and that are separated in time. Since starling songs can be much longer than the starlings' auditory memory persistence time for tonal stimuli, the birds could benefit if they had even longer memory persistence times for the complex song motifs. Further testing will be needed to find out whether the starlings' auditory memory persistence for behaviorally relevant stimuli is improved.

C. Effect of frequency difference between sample and test stimulus

We also looked at the effect of the salience of the difference between sample and test on short-term memory performance. Figure 4 shows the hit rate for sample-test combinations with frequency differences smaller than one octave and larger than or equal to one octave as a function of the delays presented. Performance was significantly worse for

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Zokoll *et al.*: Auditory memory persistence in a songbird   2849

frequency differences smaller than one octave than for frequency differences greater or equal to one octave. The starlings' FDL at 1 and 4 kHz indicate that these results are not caused by limitations of frequency resolution (Langemann and Klump, 1992: FDL of about 12 Hz at 1 kHz; and 27 Hz at 4 kHz). The minimum frequency difference of one-third octave applied in the present study in starlings is well above their FDL possibly providing more salient differences than those applied in the human studies mentioned above using tones differing in frequency by little more than the FDL. The cochlear patterns of excitation for the set of tones of different frequencies used in the present study, however, broadly overlap (Buus *et al.*, 1995). At a level of about 60 dB the excitation pattern extends by about one octave above and below the tone frequency. Thus, excitation patterns of tones separated by less than one octave broadly overlap. This overlap could explain that tones separated by less than one octave cannot be discriminated as easily as tones with larger frequency separation in a task involving auditory memory.

Humans show a decreasing ability to distinguish between the representations of the frequency of a sample stimulus and the frequency of a test stimulus if the delay between sample and test is increased. This has been shown by means of a declining ability to discriminate pitch or frequency with increasing delay (e.g., Bachem, 1954; Harris, 1952; Koester, 1945). In a study by Bachem (1954), memory persistence was measured as the relative threshold of frequency discrimination ($\Delta F/F$), or the error with which the pitch of the standard was recognized, as a function of the delay between sample and test stimuli. Bachem (1954) found a continuing deterioration of pitch recognition up to a week. The results show that with longer delays, human subjects need a greater frequency difference between sample and test stimuli to perform equally well. Wickelgren (1966, 1969) reported an improved performance for increased frequency differences between a sample and test tone (15 vs 10 Hz, and 50, 40, 30 vs 10 Hz) separated by a tone-filled delay of up to 8 s. Semal *et al.* (1996) found a similar effect when studying small to medium, and large differences in pitch of complex harmonic tones separated by a delay of 6 s (filled with intervening sounds). Error rates were lower, and detection performance was higher for stimuli with medium or large pitch differences compared to stimuli with small pitch differences.

D. Effects of training on auditory memory persistence times

In the present study we observed large interindividual differences in auditory memory persistence time. It is unlikely that this variation reflects interindividual differences in auditory frequency resolution since all frequency differences presented in the present study were well above the FDL (Langemann and Klump, 1992). There was a difference in previous experience between the two groups of starlings showing long (16.8 and 19.7 s) and short (between 4.1 and 6.4 s) auditory memory persistence times, respectively. The two birds with the long auditory memory persistence times (Fig. 2, middle row) had been previously trained in a psychoacoustic task requiring the detection of amplitude modulations of tones (Klump *et al.*, 2001) which was completed

more than 2 years before the present study started. These two birds experienced no discrimination training in the transition between this and the previous study. The three birds with short auditory memory persistence time had no previous experience with a psychoacoustical task. We have no indication on the mechanisms by which previous experience could have affected the auditory memory persistence time.

To test whether training effects occurred within the present study that could have led to an interindividual variation of auditory memory persistence times we repeated the determination of auditory memory persistence time with a completely new set of frequencies. Only one of the three birds with short auditory memory persistence times showed a substantial improvement (Table II), while the other two birds showed no improvement. In one of the two birds with long auditory memory persistence time the performance deteriorated with the new set of frequencies while in the second bird performance reached the same level. Analyzing the data from the five birds, there was no significant difference between the auditory memory persistence times in the two experiments. Thus, the difference in auditory memory persistence time seems to reflect individual differences in auditory memory rather than the status of training with a certain set of frequencies. This is similar to previous results of studies on humans in which the effect of practice on the performance was also shown to be negligible in delayed pitch discrimination tasks (Semal and Demany, 1991; Koester, 1945).

E. Conclusion

Since starlings need to assess the complexity of songs composed of repeated elements when evaluating the quality of mating partners or rivals they have to rely on extended auditory short-term memory. The results of this study show that the European starling is a promising model for studying auditory short-term memory. Similar to what is known from human observers, the birds' performance deteriorated with increasing delay before the test stimulus. Furthermore the performance of the starlings was influenced by the salience and number of repetitions of the stimuli.

Adret-Hausberger, M., Güttinger, H.-R., and Merkel, F. W. (**1990**). "Individual life history and song repertoire changes in a colony of starlings (*Sturnus vulgaris*)," Ethology **84**, 265–280.
Adret-Hausberger, M., and Jenkins, P. F. (**1988**). "Complex organization of the warbling song in the European starling Sturnus vulgaris," Behaviour **107**, 138–156.

Zokoll *et al.*: Auditory memory persistence in a songbird

Bachem, A. (**1954**). "Time factors in relative and absolute pitch determination," J. Acoust. Soc. Am. **26**, 751–753.

Baddeley, A. (**2003**). "Working memory: Looking back and looking forward," Nat. Rev. Neurosci. **4**, 829–839.

Böttcher-Gandor, C., and Ullsperger, P. (**1992**). "Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval," Psychophysiology **29**, 546–550.

Brainard, M. S., and Doupe, A. J. (**2002**). "What songbirds teach us about learning," Nature (London) **417**, 351–358.

Buus, S., Klump, G. M., Gleich, O., and Langemann, U. (**1995**). "An excitation-pattern model for the starling (*Sturnus vulgaris*)," J. Acoust. Soc. Am. **98**, 112–124.

Clément, S., Demany, L., and Semal, C. (**1999**). "Memory for pitch versus memory for loudness," J. Acoust. Soc. Am. **106**, 2805–2811.

Colombo, M., and D'Amato, M. R. (**1986**). "A comparison of visual and auditory short-term memory in monkeys (*Cebus apella*)," J. Exp. Psychol. **38B**, 425–448.

Cowan, N. (**1984**). "On short and long auditory stores," Psychol. Bull. **96**, 341–370.

Cowan, N. (**1997**). *Attention and Memory: An Integrated Framework*, Oxford Psychology Series No. 26 (Oxford University Press, New York).

Cowan, N., Saults, J. S., and Nugent, L. D. (**1997**). "The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone-pitch comparisons," Psychon. Bull. Rev. **4**, 393–397.

Deutsch, D. (**1970**). "Dislocation of tones in a musical sequence: A memory illusion," Nature (London) **226**, 286.

Deutsch, D. (**1972a**). "Mapping of interactions in the pitch memory store," Science **175**, 1020–1022.

Deutsch, D. (**1972b**). "Effect of repetition of standard and comparison tones on recognition memory for pitch," J. Exp. Psychol. **93**, 156–162.

Deutsch, D. (**1975**). "The organization of short-term memory for a single acoustic attribute," in *Short Term Memory*, edited by D. Deutsch and J. A. Deutsch (Academic, New York), pp. 107–152.

Dooling, R. (**2004**). "Audition: Can birds hear everything they sing?," in *Nature's Music. The Science of Birdsong*, edited by P. Marler and H. Slabbekoorn (Elsevier Academic, San Diego), pp. 206–225.

Doupe, A. J., and Kuhl, P. K. (**1999**). "Birdsong and human speech: Common themes and mechanisms," Annu. Rev. Neurosci. **22**, 567–631.

Downing, J. D., Okanoya, K., and Dooling, R. J. (**1988**). "Auditory short-term memory in the budgerigar (*Melopsittacus undulatus*)," Anim. Learn Behav. **16**, 153–156.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Eens, M., Pinxten, R., and Verheyen, R. F. (**1991**). "Male song as a cue for mate choice in the European Starling," Behaviour **116**, 210–238.

Gentner, T. Q., Hulse, S. H., and Ball, G. F. (**2004**). "Functional differences in forebrain auditory regions during learned vocal recognition in songbirds," J. Comp. Physiol. A **190**, 1001–1010.

Gentner, T. Q., and Margoliash, D. (**2003**). "Neuronal populations and single cells representing learned auditory objects," Nature (London) **424**, 669–674.

Gottlieb, Y., Vaadia, E., and Abeles, M. (**1988**). "Single unit activity in the auditory cortex of a monkey performing a short term memory task," Exp. Brain Res. **74**, 139–148.

Harris, J. D. (**1952**). "The decline of pitch discrimination with time," J. Exp. Psychol. **43**, 96–99.

Herman, L. M., and Gordon, J. A. (**1974**). "Auditory delayed matching in the bottlenose dolphin," J. Exp. Anal Behav. **21**, 19–26.

Keller, T. A., Cowan, N., and Saults, J. S. (**1995**). "Can auditory memory for tone pitch be rehearsed?," J. Exp. Psychol. Learn. Mem. Cogn. **21**, 635–645.

Kidd, G., Mason, C. R., and Hanna, T. E. (**1988**). "Evidence for sensory-trace comparison in spectral shape discrimination," J. Acoust. Soc. Am. **84**, 144–149.

Klump, G. M., Langemann, U., Friebe, A., and Hamann, I. (**2001**). "An animal model for studying across-channel processes: CMR and MDI in the European starling," in *Proceedings of the 12th International Symposium on Hearing: Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 266–272.

Koester, T. (**1945**). "*The time error and sensitivity in pitch and loudness discrimination as a function of time interval and stimulus level*," *Archives of Psychology*, Vol. **297**, edited by R. S. Woodworth (Columbia University, New York).

Konorski, J. (**1959**). "A new method of physiological investigation of recent memory in animals," Bull. Acad. Pol. Sci. Cl. VI. **7**, 115–117.

Kowalska, D. M. (**1997**). "The method of training dogs in auditory recognition memory tasks with trial-unique stimuli," Acta Neurobiol. Exp. (Warsz) **57**, 345–352.

Kowalska, D. M., Kusmierek, P., Kosmal, A., and Mishkin, M. (**2001**). "Neither perirhinal/entorhinal nor hippocampal lesions impair short-term auditory recognition memory in dogs," Neuroscience **104**, 965–978.

Langemann, U., Hamann, I., and Friebe, A. (**1999**). "A behavioral test of presbycusis in the bird auditory system," Hear. Res. **137**, 68–76.

Langemann, U., and Klump, G. M. (**1992**). "Frequency discrimination in the European starling (*Sturnus vulgaris*): A comparison of different measures," Hear. Res. **63**, 43–51.

National Institute of Mental Health (**2002**). Methods and Welfare Considerations in Behavioral Research with Animals: Report of a National Institutes of Health Workshop. A. R. Morrison, H. L. Evans, N. A. Ator, and R. K. Nakamura (eds). NIH Publication No. 02-5083 (U.S. Government Printing Office, Washington, DC).

McCarthy, D. C., and Davison, M. (**1991**). "The interaction between stimulus and reinforcer control on remembering," J. Exp. Anal Behav. **56**, 51–66.

Moore, B. C. J., and Glasberg, B. R. (**1989**). "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation," J. Acoust. Soc. Am. **86**, 1722–1732.

Mountjoy, D. J., and Lemon, R. E. (**1996**). "Female choice for complex song in the European starling: A field experiment," Behav. Ecol. Sociobiol. **38**, 65–71.

Pasternak, T., and Greenlee, M. W. (**2005**). "Working memory in primate sensory systems," Nature Reviews **6**, 97–107.

Paule, M. G., Bushnell, P. J., Maurissen, J. P. J., Wenger, G. R., Buccafusco, J. J., Chelonis, J. J., and Elliott, R. (**1998**). "Symposium overview: The use of delayed matching-to-sample procedures in studies of short-term memory in animals and humans," Neurotoxicol. Teratol. **20**, 493–502.

Peterson, L. R. (**1982**). "This week's citation classic: Peterson, L R & Peterson, M J. Short-term retention of individual verbal items. J. Exp. Psychol. 58: 193-198, 1959," Curr. Contents, Soc. Behav. Sci. **36**, 22.

Peterson, L. R., and Peterson, M. J. (**1959**). "Short-term retention of individual verbal items," J. Exp. Psychol. **58**, 193–198.

Sams, M., Hari, R., Rif, J., and Knuutila, J. (**1993**). "The human auditory sensory memory trace persists about 10 sec: Neuromagnetic evidence," J. Cogn Neurosci. **5**, 363–370.

Sargisson, R. J., and White, K. G. (**2003**). "The effect of reinforcer delays on the form of the forgetting function," J. Exp. Anal Behav. **80**, 77–94.

Schröger, E. (**1997**). "On the detection of auditory deviations: A pre-attentive activation model," Psychophysiology **34**, 245–257.

Schröger, E., and Berti, S. (**2000**). "Distracting working memory by automatic deviance-detection in audition and vision," in *Working on Working Memory*, edited by E. Schröger, A. Mecklinger, and A. D. Friederici, Leipzig Series in Cognitive Sciences Vol. 1 (Leipziger Universitätsverlag, Leipzig), p. 5.

Semal, C., and Demany, L. (**1991**). "Dissociation of pitch from timbre in auditory short-term memory," J. Acoust. Soc. Am. **89**, 2404–2410.

Semal, C., Demany, L., Ueda, K., and Hallé, P.-A. (**1996**). "Speech versus nonspeech in pitch memory," J. Acoust. Soc. Am. **100**, 1132–1140.

Staddon, J. E. R. (**1983**). *Adaptive Behavior and Learning* (Cambridge University Press, Cambridge).

Weavers, R., Foster, T. M., and Temple, W. (**1998**). "Reinforcer efficacy in a delayed matching-to-sample task," J. Exp. Anal Behav. **69**, 77–85.

White, K. G. (**2001**). "Forgetting functions," Anim. Learn Behav. **29**, 193–207.

White, K. G., and Wixted, J. T. (**1999**). "Psychophysics of remembering," J. Exp. Anal Behav. **71**, 91–113.

Wickelgren, W. (**1966**). "Consolidation and retroactive interference in short-term recognition memory for pitch," J. Exp. Psychol. **72**, 250–259.

Wickelgren, W. (**1969**). "Associative strength theory of recognition memory for pitch," J. Math. Psychol. **6**, 13–61.

Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.

Wright, A. A., Shyan, M. R., and Jitsumori, M. (**1990**). "Auditory same/different concept learning in monkeys," Anim. Learn Behav. **18**, 287–294.

Yoo, S.-S., Lee, C. U., and Choi, B. G. (**2001**). "Human brain mapping of auditory imagery: Event-related functional MRI study," NeuroReport **12**, 3045–3049.

Directional loudness and binaural summation for wideband and reverberant sounds[a)]

Ville Pekka Sivonen[b)]
Department of Otorhinolaryngology, University Hospital of Oulu, P.O. Box 22, 90029 OYS, Finland

In an earlier investigation [Sivonen and Ellermeier, J. Acoust. Soc. Am. **119**, 2965–2980 (2006)], the effect of sound incidence angle on loudness was investigated for anechoic, narrowband sounds. In the present follow-up investigation, the effect of incidence angle on loudness was investigated using wideband sounds under anechoic conditions and narrowband sounds under reverberant conditions. Five listeners matched the loudness of a sound coming from five incidence angles in the horizontal plane to that of the same sound with frontal incidence. These directional loudness matches were obtained with an adaptive, two-alternative, two-interval, forced-choice procedure. The stimuli were presented to the listeners via individual binaural synthesis. The results show that loudness depends on sound incidence angle in both experiments. The wideband and reverberant sounds, however, yielded significantly smaller directional effects than had been obtained for the same listeners when anechoic, narrowband sounds were used. When modeling the binaural summation underlying the loudness matches, a power summation of the at-ear signals yielded good predictions for all types of stimuli investigated. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717497]

I. INTRODUCTION

Despite the vast amount of research on spatial hearing, little is known about how the location of a sound source affects its perceived loudness. The research on binaural and spatial hearing has largely focused on human sound localization (Blauert, 1997). In the classical research on loudness (Zwicker and Fastl, 1999), however, the emphasis has been on temporal and spectral aspects, while the spatial aspects have largely been overlooked.

These were explicitly addressed in a recent series of studies determining the effect of sound incidence angle on loudness (Sivonen *et al.*, 2005; Sivonen and Ellermeier, 2006a). In these investigations, anechoic, narrowband stimuli were used for obtaining loudness matches between a frontal reference sound source and a comparison source located at various incidence angles in the horizontal or the median plane. These *directional loudness matches* showed considerable dependence on the incidence angle, the center frequency of the stimuli, and the individual listener.

In Sivonen and Ellermeier (2006a), individual head-related transfer functions (HRTFs) were measured for each incidence angle, in order to relate the changes in the actual at-ear exposures to the observed changes in loudness as a function of sound incidence. The individual at-ear exposures were then utilized in modeling the *binaural summation* underlying the directional loudness matches, i.e., it was determined how the signals at the two ears should be summed to

result in a single binaural loudness percept. Anechoic, narrowband stimuli were also used in an earlier study (Robinson and Whittle, 1960) for obtaining directional loudness matches and for modeling binaural summation, but only mean data were reported as opposed to the individual analyses attempted by Sivonen and Ellermeier (2006a).

A. Acoustic exposure and directional loudness

While the use of anechoic, narrowband stimuli (Robinson and Whittle, 1960; Sivonen *et al.*, 2005; Sivonen and Ellermeier, 2006a) may be expected to maximize frequency-specific directional effects on loudness, such stimuli lack ecological validity (e.g., Neuhoff, 2004). Note that most sounds in our environment are wideband and, in addition to the direct sound from a source to our ears, include reverberation. It is thus unclear whether the large variation (of up to 10 dB) in the directional loudness matches observed for the narrowband stimuli (Sivonen *et al.*, 2005; Sivonen and Ellermeier, 2006a) would also occur for stimuli resembling sounds our ears encounter in real life.

Based on measurements of HRTFs (see, Møller *et al.*, 1995), the directional variation in anechoic at-ear exposures shows a considerable frequency dependence. At low frequencies, the human head and torso are relatively small compared to the wavelength. Thus, the directional differences in the HRTF magnitude spectra are relatively small. At high frequencies, the magnitude spectra are characterized by many closely spaced dips, the spectral location of the dips varying with the angle of incidence. Therefore, the directional differences relevant for loudness processing may to some extent be averaged out across a wider bandwidth.

In contrast to the direct sound from a source to our ears, reverberation in sound fields consists of sound waves which

---

have been reflected a number of times from physical obstacles located in space. The incidence angles of the reflections are, by and large, different from the incidence angle of the direct sound. In addition, the more time that passes from the arrival of the direct sound, the more diffuse the reverberation becomes. In an ideal diffuse field, where the sound reaches the listener's ears with equal (acoustic) intensity from all directions, the orientation of the listener does not affect the at-ear exposures.

Therefore, stimulation which is closer to real-life sounds than the anechoic, narrowband stimuli, should either be wider in bandwidth or include reverberation. This, however, may result in relatively small directional changes in at-ear exposures, and thus smaller directional effects on loudness may be expected than for the anechoic, narrowband stimuli used earlier (Robinson and Whittle, 1960; Sivonen *et al.*, 2005; Sivonen and Ellermeier, 2006a).

B. Earlier work

Anechoic, wideband sounds have been used for obtaining directional loudness matches by Remmers and Prante (1991) and Jørgensen (2002), where the matches varied by up to 3 dB as a function of incidence angle. In an investigation of multichannel audio reproduction by Bech (1998), a reverberant listening room was used, and a directional effect of approximately 1 dB for wideband stimuli was obtained, although this was investigated only over a limited number of angles.

In Bech (1998) and Jørgensen (2002), the loudness matches were reported only as relative level adjustments for the sound sources at various incidence angles. The results are thus not generalizable, since the actual at-ear exposures underlying the loudness matches were not considered.

In Remmers and Prante (1991), an attempt was made to predict the directional loudness matches using generic, artificial-head measurements. From the research on HRTFs, large individual differences in at-ear exposures for various incidences may be expected, and thus, relating mean loudness data to artificial-head predictions, as done in Remmers and Prante (1991), might wash out some of the detail in the results. Artificial-head measurements were also utilized in Tuomi and Zacharov (2000) for simulating the effect of incidence angle on binaural loudness perception. However, subjective data, supporting the simulations, were not reported by the authors.

Relating at-ear exposures to directional loudness data is important, not only for the application of the level-alignment of multichannel sound-reproduction systems, but also for predicting loudness from acoustical measurements made with an artificial head, which are widely used. Thus, in order to generalize the findings on the binaural loudness summation of anechoic, narrowband sounds (Robinson and Whittle, 1960; Sivonen and Ellermeier, 2006a), a wider range of stimuli should be utilized in investigating the effect of sound incidence angle on loudness.

C. Present investigation

The aim of the present study is to investigate the effect of incidence angle on loudness both for wideband and reverberant sounds. In addition to obtaining directional loudness matches, *individual*, listener-specific at-ear exposures are related to the loudness matches, as was done in the earlier experiment (Sivonen and Ellermeier, 2006a) using anechoic, narrowband stimuli. The present study consists of two parts: Experiment 1, in which the effect of sound incidence angle on loudness is investigated for anechoic, *wideband* sounds, and Experiment 2, where the stimuli are *reverberant* but narrow in bandwidth.

II. METHOD

A. Subjects

Five listeners (four male, between 26 and 48 years of age), all of whom had participated in previous experiments on directional loudness (Sivonen *et al.*, 2005; Sivonen and Ellermeier, 2006a), took part in the present study. The listeners' hearing thresholds were determined using pure-tone audiometry with octave steps between 0.25 and 8 kHz. None of the hearing thresholds were greater than 15 dB hearing level re the standard values specified in ISO 389-1 (1998).

B. Apparatus

1. Listening tests

Stimuli were played back to the listeners over headphones (Beyerdynamic DT-990) using individual binaural synthesis, as in Sivonen *et al.* (2005). Since the binaural synthesis was static, the listeners were seated in a chair with a head rest to restrict head movements. For each listener, the chair was adjusted so that the center of the listener's head was aligned with the center position of the test setup.

The listeners' responses were collected using a two-button response box with light-emitting diodes above each button indicating observation intervals. A model of this box (with larger lights) was placed in front of the listeners to prevent them from tilting their heads toward the response box in their hands. The rest of the listening-test setup consisted of a computer (PC) with a high-quality sound card (RME DIGI96/8 PST), a custom-made programmable attenuator, and a power amplifier (Pioneer A-616) combined with a passive attenuator for unity gain, all placed in a control room [for details of the apparatus, see Sivonen *et al.* (2005)].

In order to improve the plausibility of the binaural synthesis, loudspeakers at the intended incidence angles were visible to the listeners, even though the stimuli were actually played back over headphones. The distance from each loudspeaker to the center position of the setup was 2.0 m. Experiment 1 took place in an anechoic chamber, whereas a reverberant listening room, conforming to IEC 60268-13 (1998), was used for Experiment 2.

2. Measurements

In order to prepare stimuli for the individual binaural synthesis, a number of measurements were carried out. Indi-

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Ville Pekka Sivonen: Directional loudness   2853

vidual (anechoic) HRTFs and headphone transfer functions (PTFs), measured in connection with the earlier directional-loudness investigation (Sivonen and Ellermeier, 2006a), were used to synthesize stimuli for Experiment 1.

For Experiment 2, individual binaural room impulse responses (BRIRs) and PTFs were measured in a listening room, the dimensions of which were $2.8 \times 4.2 \times 7.8$ m ($H \times W \times L$). A loudspeaker (Vifa M10MD-39, mounted in a hard plastic ball) at a distance of 2.0 m from the measurement position was used as a sound source. The loudspeaker and the measurement position were at a height of 1.3 m and at a distance of at least 1.6 m from the closest wall, positioned nonsymmetrically with the dimensions of the room. The BRIRs were measured with two miniature microphones (Sennheiser KE-4-211-2, one specifically for each ear), which were inserted in earplugs (E·A·R Classic, halved in length) and mounted flush with the ear-canal entrance. A pressure-field microphone (Brüel & Kjær Type 4133) was used as a reference microphone for measuring the loudspeaker's anechoic response and its response in the room, and determining the responses of the two miniature microphones.

For the BRIR measurements, the position of the loudspeaker was fixed in the room, and the listeners were rotated at the measurement position in order to obtain various incidence angles. The listeners were seated in the same chair as used in the listening test and the listeners' position was aligned with a fixed laser, pointing to the center of the setup from 90° incidence. The orientation of each listener was calibrated so that for frontal incidence, the chair was rotated until time-aligned impulse responses were obtained at the two ears. Then, a polar-coordinate graph was attached to the top of the listener's head, denoting 0° incidence on the monitor of a video camera above the center position of the setup. Reference points for given listener orientations were marked on the video monitor, and the chair was rotated to obtain the desired incidence angle.

In addition to the PC, the sound card, and the power amplifier of the listening-test setup, the equipment for the BRIR measurements included an audio interface for AD/DA conversion (RME ADI DS-8) and a custom-made preamplifier for the miniature microphones. The measurements were made using a maximum-length sequence with an order of 15 and a sampling rate of 48 kHz, resulting in an impulse response with duration 0.68 s. All microphone signals were bandpass filtered between 0.02 and 22.5 kHz and appropriately amplified by measurement amplifiers (Brüel & Kjær Type 2607, Type 2236, or Type 2690 Nexus).

C. Stimuli

Before processing stimuli for the binaural synthesis, all sound signals were 1 s long, including 20-ms raised-cosine rise and fall ramps at the beginning and at the end of each signal. Stimuli were generated for six different incidence angles: 0°, 30°, 60°, 90°, 135°, and 180° of azimuth, in the left hemisphere of the horizontal plane with 0° elevation.

In both Experiments 1 and 2, the stimuli were such, that if measured with a monophonic microphone in the absence of a listener, the same sound pressure would be obtained for

TABLE I. Room-acoustics parameters (ISO 3382, 1997) for the standard listening room for the specific source and receiver positions, measured using third-octave bands. The early decay time (EDT) is obtained from the initial 10 dB and reverberation time (RT) from the portion between -5 and -35 dB of the decay curve, both calculated as the time (in seconds) required for a decay of 60 dB. C_{50} and C_3 are the balance between early/late and direct/reverberant energy (as a ratio in decibels), before and after 50 and 3 ms, respectively.

f_c(kHz)	EDT(s)	RT(s)	C_{50}(dB)	C_3(dB)
1	0.37	0.44	9.5	−10.5
5	0.41	0.62	10.4	0.9

all incidence angles. Thus, if this monophonic pressure was used to make loudness predictions, loudness would be constant as a function of incidence angle. Note, however, that when emanating from different incidence angles, the stimuli resulted in directional dependencies of at-ear exposures. In the analysis of the listening test results, these dependencies were determined for each individual listener.

1. Experiment 1: Wideband noise

A pink-noise signal, bandpass filtered between 0.1 and 10 kHz, was used as a wideband stimulus. It was played back at an overall level corresponding to 55 dB sound pressure level (SPL) measured in the free field. At this overall level, the loudness of the pink noise for frontal incidence was roughly the same as that of the 65-dB-SPL narrowband noises used in the earlier studies (Sivonen et al., 2005; Sivonen and Ellermeier, 2006a), based on predictions made from the loudness model of Moore et al. (1997).

For the binaural synthesis, the pink noise was convolved with individual HRTFs for the incidence angles. Furthermore, individual equalization filters for the headphones were applied [for details concerning the equalization, see Sivonen et al. (2005)].

2. Experiment 2: Reverberant stimuli

Third-octave noise bands, centered at 1 and 5 kHz, were used as reverberant stimuli. These narrowband stimuli were synthesized at an overall level corresponding to 65 dB SPL measured in the free field, i.e., without the effect of reverberation.

The effect of the listening room in the *absence* of a listener can be inferred from measurements with the monophonic reference microphone, which were used in deriving some room-acoustics parameters according to ISO 3382 (1997). In Table I, these parameters are shown for third-octave bands at the two center frequencies used in the listening test. Both the early decay time (EDT) and reverberation time (RT) show that the listening room is more reverberant at 5 kHz than at 1 kHz. The ratio between early and late energy (C_{50}) in decibels indicates that most of the energy arrives at the listening position within the first 50 ms, C_{50} being fairly similar at the two center frequencies. However, when looking at the ratio between direct and reverberant energy (in the present study, before and after 3 ms from the direct sound), at 1 kHz most of the energy is in the reverberation, whereas the ratio is close to unity at 5 kHz. This can be explained by

the increased directivity of the loudspeaker[1] at 5 kHz, i.e., the reverberation of the room is excited less at the higher center frequency.

In addition to the listening room, each listener's physique had idiosyncratic effects on the at-ear exposure levels. These are captured by the individual BRIRs (for details, see Sivonen and Ellermeier, 2006b). BRIRs for all listeners and for the six incidence angles were obtained by rotating the listener in the chair, carrying out measurements at each incidence. The anechoic, on-axis response of the loudspeaker, measured at the same distance as the BRIRs in the absence of a listener, was compensated for in the BRIR computations. In this manner, changes in the direct sound comparable to HRTFs were obtained, while the effect of the reverberant energy was fairly constant[2] as a function of incidence angle. In the experiment proper, the virtual acoustics of the room were "rotated" around the listener. This went largely unnoticed, because static sound sources were synthesized and the listeners could see the (dummy) loudspeakers at the intended angles.

Finally, the narrowband stimuli were convolved with the individual BRIRs, and equalization filters for the headphones were applied, based on individual PTFs measured in connection with the BRIR measurements.

D. Procedure

To make the present results comparable with those of earlier studies using the same listeners, the listening-test procedure was largely as reported in Sivonen *et al.* (2005) and Sivonen and Ellermeier (2006a).

The listeners' task was to compare the loudness of a sound with a given frequency content (Experiment 1: pink noise, Experiment 2: 1 or 5 kHz narrowband noise) synthesized with one of the comparison incidence angles (30°, 60°, 90°, 135°, or 180° of azimuth) to the loudness of the same sound with frontal incidence, i.e., 0° azimuth. Loudness matches with the frontal reference were obtained in an adaptive, two-interval, two-alternative, forced-choice procedure.

On each trial, the listeners were played a pair of sounds, the frontal reference and a comparison in a randomized order, and they had to judge which of the two sounds was louder. If the comparison was judged louder, its level was decreased by a given amount, whereas if the reference was judged louder the level of the comparison was increased by the same amount. The initial step size of such an adaptive track was 4 dB before the second reversal in the listener's judgments and 1 dB for the rest of the track.

In one adaptive track, eight reversals were obtained and the mean of the levels at the last six was taken as the loudness match. Eight replications of each factorial combination of stimulus frequency and incidence angle were collected for each listener. Half of the replications started with the comparison sound at +10 dB and the other half with it at −10 dB relative to the frontal reference. In Experiment 2, the listeners also had to match the frontal reference to itself, as an indication of the baseline variability of the loudness matches.

The order of the adaptive tracks (stimulus frequency, incidence angle, and starting level) was randomized within
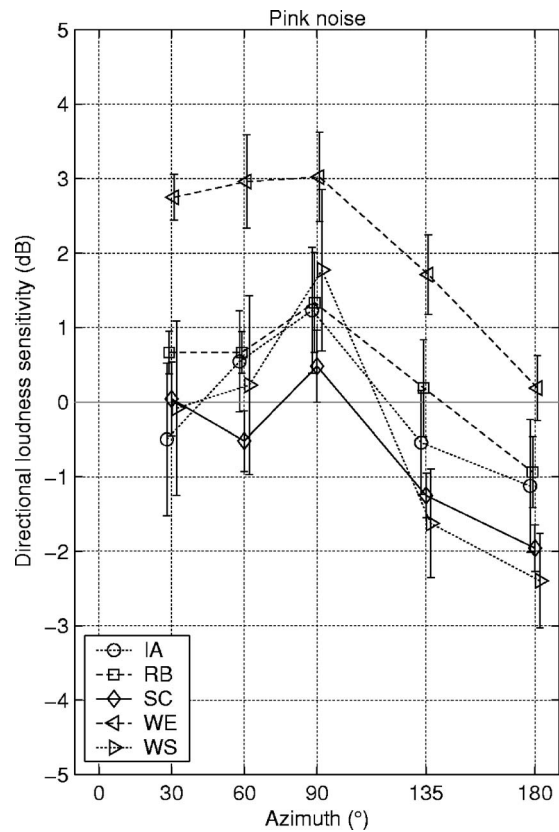
FIG. 1. Results of Experiment 1: Individual directional loudness sensitivities (in decibels) ±95% confidence intervals in the left hemisphere of the horizontal plane for anechoic pink-noise stimuli, obtained from five listeners.

one set of replications. The tracks were divided into blocks of eight and interleaved within a block, one 1-h listening session consisting of a maximum of four blocks, each lasting approximately 10 min. All listeners participated in Experiment 1 first and then in Experiment 2. The total duration was 6 h, including the BRIR- and PTF-measurement sessions.

III. RESULTS

In both Experiments 1 and 2, loudness matches with the frontal reference were obtained for five incidence angles in the horizontal plane. Thus, the raw data are the relative sound pressure levels each synthesized sound source would have to be set to, in order to be perceived as equally loud as the reference. The inverses of these gain adjustments were termed *directional loudness sensitivities*, in agreement with earlier studies, a positive value indicating a direction from which the stimulus is perceived as louder than the frontal reference.

A. Experiment 1: Wideband noise

Individual directional loudness sensitivities for the wideband, pink-noise stimulus are plotted in Fig. 1 for all five listeners. The data show that loudness is not constant as a function of incidence angle: For 16 out of 25 data points in Fig 1, the 95% confidence intervals do not reach the 0-dB line denoting the relative level of the frontal reference.
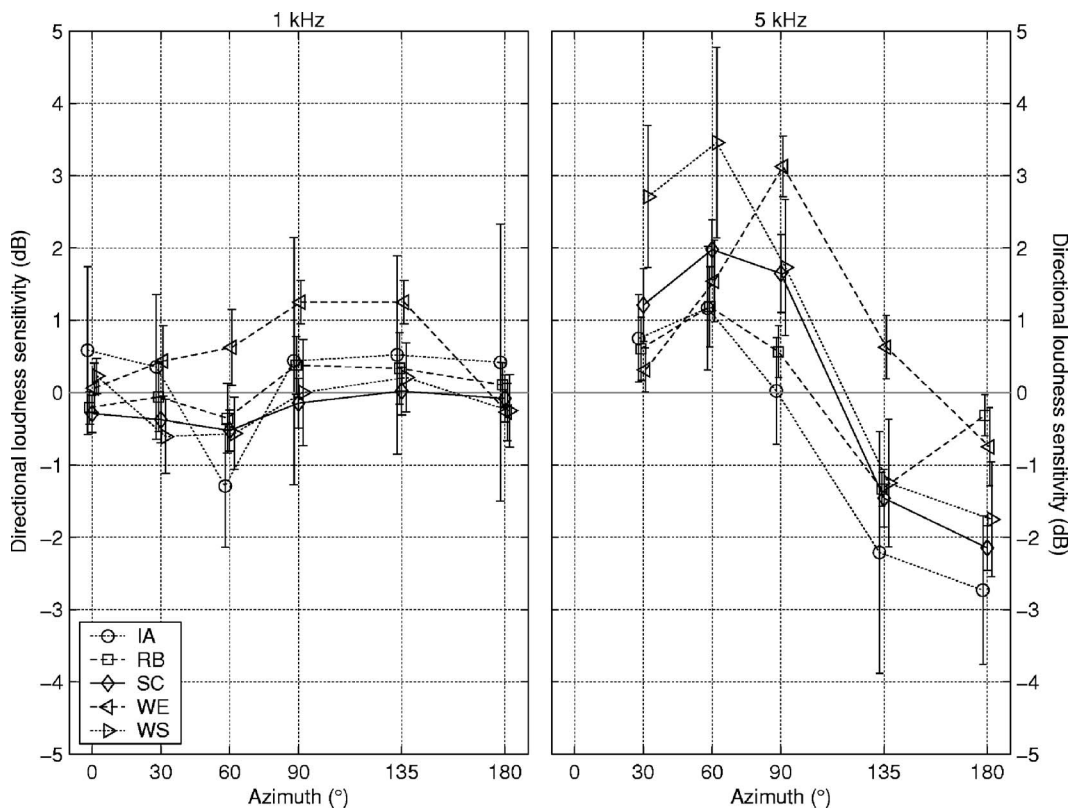
FIG. 2. Results of Experiment 2: Individual directional loudness sensitivities (in decibels) ±95% confidence intervals in the left hemisphere of the horizontal plane for reverberant narrowband stimuli, obtained from five listeners.

The directional loudness sensitivities also show individual differences. For subject WE, all data points are above the 0-dB line, the maximum value being around +3 dB for the 90° incidence, i.e., a sound source at that angle is perceived as louder and has to be attenuated by 3 dB to produce a loudness match with the frontal reference. For subject SC, however, most of the data points are either close to or below the 0-dB line, the latter implying that the level of the sound source has to be increased for a match with the reference.

To determine the significance of the observed directional loudness sensitivities, a repeated-measures analysis of variance (ANOVA) was performed with the incidence angle as a fixed factor and listener as a random factor. The ANOVA returned a significant effect of incidence angle [$F(4,16) = 26.63; p < 0.001$], confirming that loudness is dependent on the angle of incidence for the wideband stimulus. Also, the interaction between incidence angle and listener was significant [$F(16,175) = 2.59; p = 0.001$], revealing statistically significant individual differences in the pattern of the listeners' directional loudness matches.

B. Experiment 2: Reverberant stimuli

Individual directional loudness sensitivities for the reverberant narrowband stimuli at the two center frequencies are plotted in Fig. 2. The directional effect at 1 kHz is small: Most of the data lie within ±1 dB of the 0-dB line, and only for 9 out of 30 data points do the 95% confidence intervals not cover the 0-dB line. Due to the smaller spread of the individual data, differences in the patterns between listeners are less obvious than in Experiment 1 (see Fig. 1), although

again the data for subject WE display the highest directional loudness sensitivities for the comparison incidence angles. Listeners accurately matched the frontal source to itself, as seen in the 0-dB (i.e., identity) outcome for the left-most data point in Fig. 2.

At 5 kHz the data vary over a larger range of approximately 6 dB, and in all but one case the 95% confidence intervals do not cover the 0-dB line. The individual differences are also larger in magnitude, even though for all subjects except WE, the shapes of the curves as a function of incidence angle are quite similar.

A two-factor (incidence angle × center frequency) ANOVA on the data of Experiment 2 showed a significant interaction between incidence angle and center frequency [$F(4,16) = 15.78; p < 0.001$]. This implies that the shapes of the directional loudness-sensitivity curves are different at the two center frequencies. Furthermore, the interaction between incidence angle, center frequency, and listener [$F(16,350) = 4.95; p < 0.001$] revealed statistically significant individual differences in the pattern of the matches for the reverberant stimuli, as for the wideband stimuli of Experiment 1.

IV. DISCUSSION

Sound incidence angle had an effect on loudness both for the anechoic, wideband noise of Experiment 1 and the reverberant, narrowband stimuli of Experiment 2. The directional loudness sensitivities derived from the loudness matches also exhibited considerable dependencies on the individual listener.

The sound sources in the present study were synthesized with a flat frequency response, and calibrated so that the same sound pressure would be measured in the center position of the listening-test setup for all incidence angles. A monophonic measurement [p_1, in the absence of a listener, see Møller *et al.* (1995)] would thus fail to predict the observed directional loudness matches.

Therefore, the directional changes in the at-ear exposures were analyzed and related to the directional loudness sensitivities, as had been done in the earlier directional-loudness investigation (Sivonen and Ellermeier, 2006a). In the present study, the HRTFs and BRIRs, as measured at the blocked entrance to ear canal [p_2, see Møller *et al.* (1995)], were used in the stimulation. These measurements of individual HRTFs and BRIRs were utilized in determining the exposures at each listener's ears for Experiment 1 with anechoic wideband stimuli and for Experiment 2 with reverberant narrowband stimuli, respectively.

A. Experiment 1

1. Considerations regarding stimulus properties

The spectral bandwidth of a stimulus of fixed overall level has been shown to affect perceived loudness, a phenomenon which is known as the spectral summation of loudness (Zwicker and Fastl, 1999). Loudness models (e.g., Moore *et al.*, 1997) have been developed to account for this summation of loudness across critical bands. Since wideband stimuli were used in Experiment 1, the spectral summation of loudness cannot be disregarded in the analysis. Therefore, the directional changes in the at-ear exposures must be analyzed as loudness values in sones, predicted by a loudness model.

Furthermore, for each subjective data point in Figs. 1 and 2, two at-ear exposure values, one for each ear, were determined. Thus the binaural loudness summation of the left- and right-ear exposures to a single binaural percept must also be considered when predicting the subjective data.

2. Modeling

The loudness model of Moore *et al.* (1997) was utilized in the present analysis with three predictions, the main difference between predictions being the emphasis given for the left- and right-ear inputs for computing binaural loudness:

The first prediction is based on an assumption that binaural loudness is determined solely by the "louder" ear, i.e., for each incidence angle by the ear, whose input sound pressure yields the higher sone value, as calculated using the model of Moore *et al.* (1997). In this manner, it was investigated whether monaural HRTF effects could predict the obtained data.

The second prediction is a power-summation model, favored in the analysis of Sivonen and Ellermeier (2006a) for a directional sound field. In this prediction, a binaural power sum of the linear left- and right-ear HRTF magnitude spectra with a given (third-octave) frequency resolution is first determined. Then the power sum is converted to the corresponding diotic (i.e., with the same signal at both ears) stimulation by halving the summed power. Finally, this (di-

otic) signal is utilized for loudness computations. For details of the power-summation model, see Sivonen (2006).

The third prediction is a loudness-summation model, where loudness is computed separately for the left- and right-ear signals and binaural loudness determined as a sum of the left- and right-ear sone values. This manner of summation is supported by some headphone studies on binaural loudness (e.g., Marks, 1978), and is implemented in Moore *et al.* (1997) and utilized, e.g., by Tuomi and Zacharov (2000).

In order to predict loudness from at-ear exposures, the loudness model of Moore *et al.* (1997) allows substitution of the HRTF-filtering stage of the model by a measured at-ear exposure, utilizing ear-drum pressures (p_4) for loudness computations. Since the individual HRTFs (and the BRIRs) of the present study were measured at the blocked entrance to the ear canal (p_2), a direction-independent transfer from the blocked entrance to the ear-drum (mean p_4/p_2) was adopted from Fig. 13 in Hammershøi and Møller (1996). For the three binaural predictions presented earlier, individual loudness matches were then obtained by varying the p_1 pressure so that the difference in sones between the frontal reference and each comparison incidence was minimized, based on the individual HRTF effects on the left- and right-ear p_4 pressures.

3. Outcome

In Fig. 3, the three predictions are plotted along with the obtained directional loudness sensitivities (of Fig. 1) in individual panels for each listener and for the mean across listeners. As was the case for the subjective data, the predicted (relative) loudness matches were converted to directional loudness sensitivities by inverting them. Even though the predictions deviate from one another by a maximum of 4 dB, their order is the same in all panels of Fig. 3: The monaural (louder-ear) loudness prediction always yields the highest values, and the prediction based on a power sum is always above the one based on a loudness sum. The power and the loudness sum are both "pulled down" with respect to the louder-ear prediction by the loudness decrease at the "softer" ear, although to a different degree: In the power sum the ear with the higher input level is emphasized more, while in the loudness sum the ear with the lower input level has a relatively large effect on binaural loudness. A more detailed discussion of different binaural summation models can be found in Sivonen and Ellermeier (2006a).

In most cases, the power sum yields the best prediction of the obtained subjective data. For subjects IA and RB, and partly for subject WS, the agreement between the actual data and the power-sum prediction is fair. For subject WE, the max-ear loudness yields the best prediction, while for subject SC, the loudness sum is somewhat closer to the measured data. However, when inspecting the mean data obtained by averaging third-octave-band HRTFs across listeners for the three predictions and the mean directional loudness sensitivities, the power-sum prediction is clearly favored for the anechoic wideband stimuli (see the lower-right panel in Fig. 3). Note that this was also the case for anechoic narrowband stimuli in Sivonen and Ellermeier (2006a).
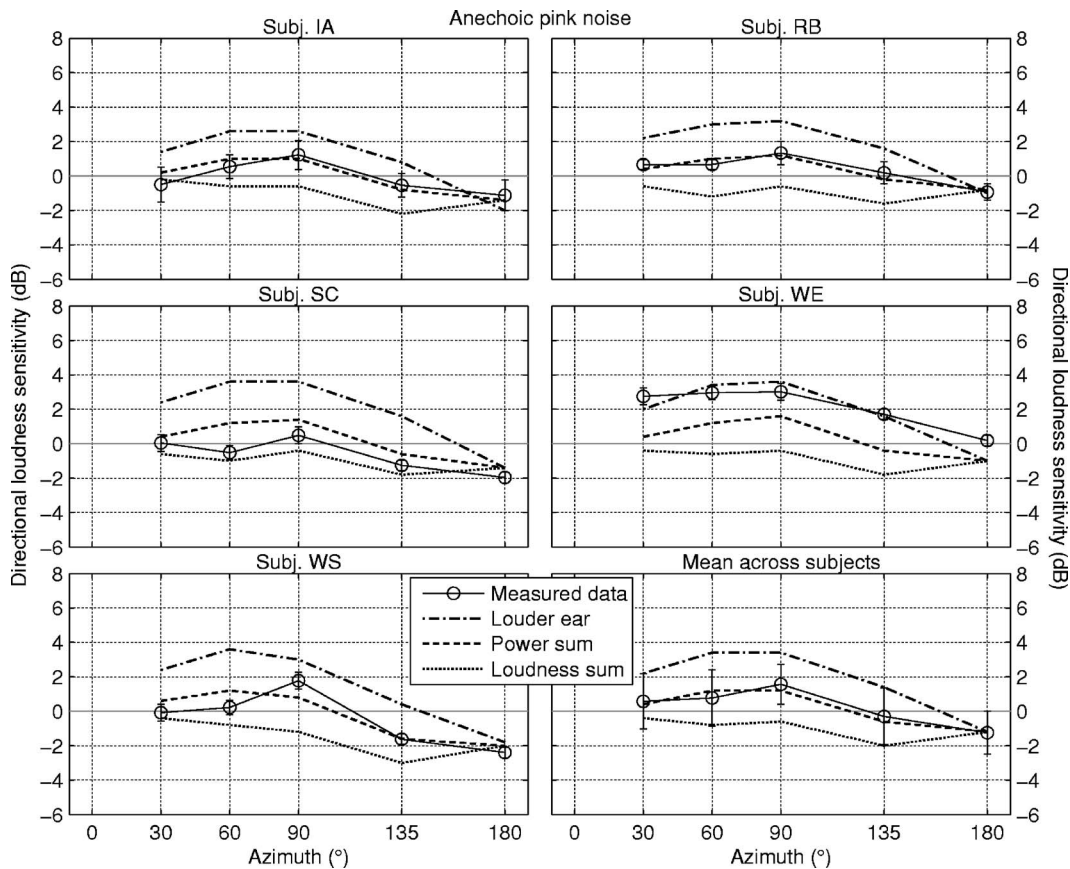
FIG. 3. Modeling Experiment 1: Directional loudness sensitivities (in decibels) ±95% confidence intervals for the anechoic wideband (pink-noise) stimuli, plotted along with predictions based on the louder ear, a binaural power sum, and a loudness sum for the five individual listeners and a mean across listeners.

B. Experiment 2

1. Considerations regarding stimulus properties

For Experiment 2, reverberant narrowband stimuli were used. The mean directional loudness sensitivities for these stimuli are plotted in Fig. 4 for comparison with data for anechoic stimuli at the same center frequencies (Sivonen *et al.*, 2005), as obtained from the same five listeners. Both the anechoic and the reverberant stimuli were presented via individual binaural synthesis, the only difference being that for the present data, the sounds included fairly constant reverberation (RT ≈ 0.5 s) as a function of incidence angle.

The directional effect was smaller for the reverberant stimuli, the mean loudness data being close to 0 dB at 1 kHz, while at 5 kHz, the effect size was roughly halved compared to the anechoic case. This discrepancy between center frequencies can be explained by the measured room-acoustics parameters in Table I: The balance between direct and reverberant energy (C_3) revealed the sound field at 1 kHz to be dominated by reverberation, while at 5 kHz the direct and reverberant energy were comparable. Thus, the changes in the direct sound, as a function of the listener's orientation relative to the sound source, were more likely to affect loudness at the higher center frequency. Furthermore, as reported in Sivonen and Ellermeier 2006a, the directional effect on loudness was larger at higher center frequencies, which can be explained by the respective changes in HRTFs. Finally, the error bars in Fig. 4 indicated slightly smaller variations between listeners for the present data, due to the

room reverberation being the same for all listeners, whereas in Sivonen *et al.* (2005) with (pure, anechoic) HRTFs the individual differences were more pronounced.
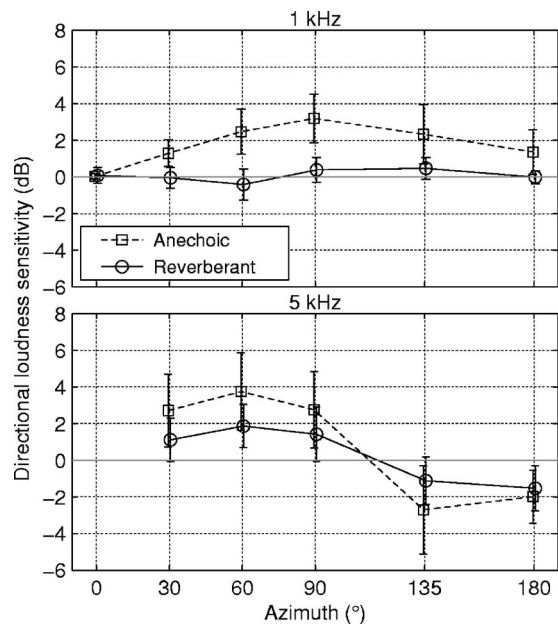
FIG. 4. Effect of reverberation on directional loudness: Mean directional loudness sensitivities ±95% confidence intervals of the means, obtained in individual binaural synthesis from five listeners for anechoic (Sivonen *et al.*, 2005) and reverberant (Experiment 2) narrowband stimuli.
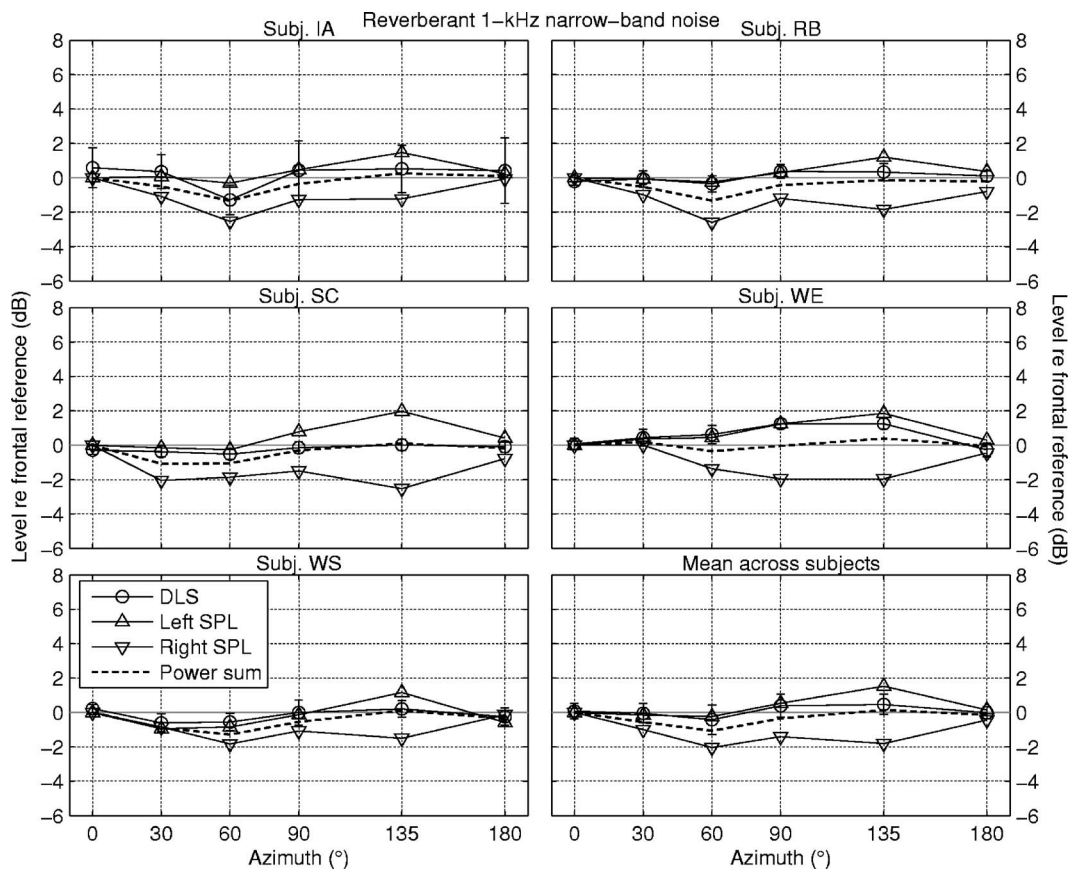
FIG. 5. Modeling Experiment 2, 1 kHz: Individual directional loudness sensitivities (DLSs) ±95% confidence intervals for five listeners, and a mean across listeners, plotted along with the effective changes in the left- and right-ear SPLs, and a prediction based on a binaural power sum.

2. Modeling

To evaluate the effect of the BRIRs on the obtained directional loudness sensitivities on an individual basis, third-octave-band, left- and right-ear sound pressure levels were computed from the stimuli convolved with the BRIRs for each listener, incidence angle, and center frequency. The at-ear SPLs were then normalized to those of the frontal reference, separately for the left and the right ears.

Then, a power-sum prediction was computed from the left- and right-ear SPLs, also normalized to the frontal reference. Note that this prediction is based on a summation of at-ear SPLs, which was fairly successful for the anechoic, narrowband stimuli used by Sivonen and Ellermeier (2006a). Unlike in the analysis of the results for the wideband stimuli, a loudness model was not utilized here.

3. Outcome

The individual directional loudness sensitivities, the relative directional changes in the at-ear SPLs, and the power-sum prediction are plotted in Figs. 5 and 6 for the 1- and 5-kHz center frequencies, respectively. Even though interaural level differences (ILDs) are present at both center frequencies, the effect of the room on the at-ear exposures is strong: In an anechoic sound field, for the same listeners and incidence angles, ILDs of over 10, and up to 30 dB were obtained by Sivonen and Ellermeier (2006a), while in the present study the corresponding ranges are 4 and 10 dB at 1 and 5 kHz, respectively. Despite the smaller variation in

ILDs, the subjective directional loudness sensitivities (with a few exceptions) lie within the bounds defined by at-ear SPLs, exhibiting a good agreement, similar to that reported for the anechoic stimuli used by Sivonen and Ellermeier (2006a).

When predicting the directional loudness data from the at-ear SPLs, differences between individuals can be observed, i.e., for some listeners (e.g., subjects IA, SC, and WS) the power sum yields a fair prediction, while for others (e.g., subject WE), the prediction is at odds with the obtained subjective data: See Figs. 5 and 6. These peculiarities, which remain even after accounting for individual differences in at-ear exposures, imply that there are differences in the underlying binaural loudness summation of the left- and right-ear signals. The mean data in both Figs. 5 and 6, however, are in good agreement with the power-sum prediction.

C. Summary

Despite the (small) individual differences between the directional loudness sensitivities and predictions based on at-ear exposures, the mean data of both the wideband and the reverberant stimuli agree well with a binaural power summation, see Figs. 3, 5, and 6. Based on the present results, it seems that for spatial sounds reaching the listener's ears with various incidences, adjustments to the modeling of the underlying binaural summation of loudness are called for.

The aim of the binaural modeling utilized here is to predict the loudness of spatial sounds from at-ear exposures.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Ville Pekka Sivonen: Directional loudness   2859
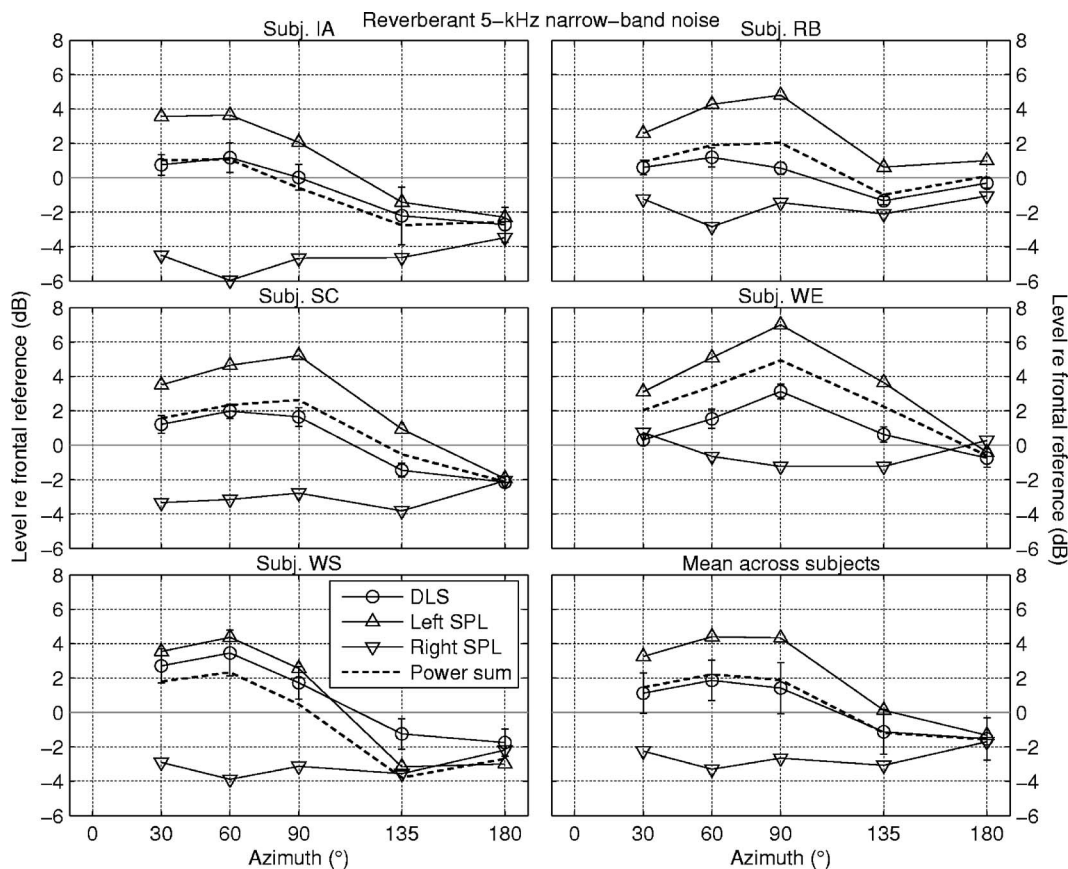
FIG. 6. Modeling Experiment 2, 5 kHz: Individual DLSs ±95% confidence intervals for five listeners, and a mean over listeners, plotted along with the effective changes in the left- and right-ear SPLs, and a prediction based on a binaural power sum.

Note that despite the fair predictions, the modeling is not based on actual processes in the auditory pathway. A more physiologically based approach has recently been taken by Moore and Glasberg (2007), modeling the effect of interaural inhibition on binaural loudness computations.

The present results show, however, that a perfect binaural summation of loudness in sones [as implemented in Moore *et al.* (1997)] does not seem to be warranted for spatial, wideband sounds. Inspecting the mean data of Fig. 3 shows that the prediction based on loudness summation is below the measured data, yielding lower directional loudness sensitivities than were obtained. As in Sivonen and Ellermeier (2006a), this prediction overestimates the effect of the (contralateral) ear having the lower-level exposure (in dB SPL) on binaural loudness. Here, the same reasoning applies, although plotting the at-ear exposure in dB SPL along with the subjective data (as in Figs. 5 and 6) is not meaningful, due to the wideband nature of the stimuli. A binaural power sum, however, gives greater emphasis to the ear having the higher-level exposure, and thus arrives at predictions above those of the loudness summation.

To sum up, computing a power sum of the HRTF magnitude spectra, converting that sum to the corresponding diotic spectrum, and only then utilizing this at-ear spectrum for loudness computations [according to the binaural power-summation model of Sivonen (2006)] gives a good prediction of the directional loudness matches for the wideband stimuli.

When the sound field is reverberant, the exposure differences between directions, and between the two ears, are smaller than those of the anechoic case, due to the diffuseness of the reverberation. Despite the reverberation in the stimuli, a quantifiable, statistically significant effect of direction on loudness was observed. Therefore, these results do not support the phenomenon of perfect loudness constancy. In a study by Zahorik and Wightman (2001), source loudness was judged to be constant for various sound-source distances in a reverberant room, irrespective of profound changes in at-ear levels. This observation of constancy could be due to instructing listeners to focus on the *source* loudness, whereas in the present series of experiments (Sivonen *et al.*, 2005; Sivonen and Ellermeier 2006a, present study), no such instructions were given.

Note that the small exposure differences between the two ears for the reverberant stimuli are not optimal for the purpose of modeling binaural loudness summation. The closer to diotic the stimulation is, the less the summation rule plays a role in predicting binaural loudness matches, since binaural summation has the same effect both for the reference and the comparison [for a more detailed discussion, see Sivonen and Ellermeier (2006a)]. In real life, the sounds we encounter are often wideband *and* reverberant, possibly yielding only small exposure differences between the ears, irrespective of the location of the source relative to the listener. Therefore, e.g., reverberant speech stimuli, selected for locations maximizing interaural exposure differences, should

be considered for future experiments on directional loudness. However, the present results for the reverberant narrowband sounds agree well with the binaural power-summation model, a model which is based on anechoic stimuli with larger exposure differences between the two ears (Sivonen and Ellermeier, 2006a).

All in all, the present results, obtained using stimuli which resemble real-life sounds more than the anechoic, narrowband stimuli used in Sivonen and Ellermeier (2006a), corroborate the validity of binaural power summation for the loudness of spatial, directional sounds.

V. CONCLUSION

The effect of sound incidence angle on loudness was investigated for wideband and reverberant sounds. The results for both types of sounds showed that loudness is not constant over incidence angles, and the directional loudness matches exhibit considerable dependence on the listener. These idiosyncrasies could, however, largely be accounted for by determining individual at-ear exposures. A binaural power-summation model yielded a fair prediction of the obtained directional loudness data, and was favored over a perfect summation of loudness in sones between the two ears, as commonly used. The present results thus suggest adjustments to the modeling of binaural loudness summation for spatial sounds.

[1] An omnidirectional sound source should be used for room-acoustics measurements. Even though this is not the case for the loudspeaker, its on-axis frequency response has roughly the same magnitude at 1 and 5 kHz.

[2] Note that when rotating the listener, the measurement positions at the ears move along the perimeter of the head. Since the loudspeaker position was fixed in the room, the changes with the measurement position in the reverberant sound were small compared to the corresponding changes in the direct sound. Further, if *assumed* that the sound field was uniform in the at-ear plane and the reverberant sound was diffuse, the rotation would not affect the reverberation at the listener's ears.

Bech, S. (**1998**). "Calibration of relative level differences of a domestic multichannel sound reproduction system," J. Audio Eng. Soc. **46**, 304–313.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.

IEC 60268-13 (**1998**). *Sound System Equipment—Part 13: Listening Tests on Loudspeakers* (International Electrotechnical Commission, Geneva, Switzerland).

ISO 3382 (**1997**). *Acoustics—Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters* (International Organization for Standardization, Geneva, Switzerland).

ISO 389-1 (**1998**). *Acoustics—Reference Zero for the Calibration of Audiometric Equipment—Part 1: Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Supra-Aural Earphones* (International Organization for Standardization, Geneva, Switzerland).

Jørgensen, S. L. (**2002**). "Matlab model for loudness perception—Directional loudness," Masters thesis, Technical University of Denmark (DTU), Copenhagen (unpublished).

Marks, L. E. (**1978**). "Binaural summation of the loudness of pure tones," J. Acoust. Soc. Am. **64**, 107–113.

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects," J. Audio Eng. Soc. **43**, 300–321.

Moore, B. C. J., and Glasberg, B. R. (**2007**). "Modeling binaural loudness," J. Acoust. Soc. Am. **121**, 1604–1612.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness," J. Audio Eng. Soc. **45**, 224–239.

Neuhoff, J. G. (**2004**). *Ecological Psychoacoustics* (Academic, London).

Remmers, H., and Prante, H. (**1991**). "Untersuchung zur Richtungsabhängigkeit der Lautstärkeempfindung von breitbandigen Schallen" ("Investigation of the directional dependence of the loudness of broadband sounds"), *Fortschritte der Akustik*—Deutsche Arbeitsgemeinschaft für Akustik (DAGA), pp. 537–540.

Robinson, D. W., and Whittle, L. S. (**1960**). "The loudness of directional sound fields," Acustica **10**, 74–80.

Sivonen, V. P. (**2006**). "Directional loudness perception—The effect of sound incidence angle on loudness and the underlying binaural summation" Ph.D. thesis, Aalborg University, Denmark (unpublished).

Sivonen, V. P., and Ellermeier, W. (**2006a**). "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation," J. Acoust. Soc. Am. **119**, 2965–2980.

Sivonen, V. P., and Ellermeier, W. (**2006b**). "Effect of direction on loudness for wideband and reverberant sounds," Proceedings of the 120th Convention of the Audio Engineering Society, Convention Paper No. 6821.

Sivonen, V. P., Minnaar, P., and Ellermeier, W. (**2005**). "Effect of direction on loudness in individual binaural synthesis," Proceedings of the 118th Convention of the Audio Engineering Society, Convention Paper No. 6512.

Tuomi, O., and Zacharov, N. (**2000**). "A real-time binaural loudness meter," presented at the 139th Meeting of the Acoustical Society of America, Atlanta, J. Acoust. Soc. Am. **107**, 2789.

Zahorik, P., and Wightman, F. L. (**2001**). "Loudness constancy with varying sound source distance," Nat. Neurosci. **4**, 78–83.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models* (Springer, New York).

Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F_0 and amplitude control[a]

Charles R. Larson[b] and Jean Sun

Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

Timothy C. Hain

Departments of Neurology, Otolaryngology, and Physical Therapy/Human Movement Sciences, Northwestern University, 645 N. Michigan, Suite 1100, Chicago, Illinois 60611

Perturbations in either voice pitch or loudness feedback lead to changes in a speaker's voice fundamental frequency (F_0) or amplitude. Voice pitch or loudness perturbations were presented individually (either pitch or loudness shift stimuli) or simultaneously (pitch combined with loudness shift stimuli) to subjects sustaining a vowel to test the hypothesis that the mechanisms for these two response types are independent. For simultaneous perturbations, pitch and loudness both changed in the same direction or in opposite directions. Results showed that subjects responded with voice F_0 or amplitude responses that opposed the direction of the respective pitch- or loudness shift stimuli. Thus, depending on the stimulus direction, both responses could either change in the same direction or in the opposite direction to each other. F_0 response magnitudes were greatest with pitch-shift only stimuli (18 cents), smallest for loudness shift stimuli (10 cents) and intermediate with pitch combined with loudness shift stimuli (13 and 16 cents). Amplitude responses were largest with +3 dB stimuli (0.96 dB) and smallest with −3 dB stimuli (0.49 dB) but were not affected by the addition of pitch-shift stimuli. Results suggest the F_0 and amplitude response mechanisms may be independent but interact in some conditions. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2715657]

I. INTRODUCTION

The importance of auditory feedback for voice control in speech and singing has been recognized for centuries, primarily from observations on the speech of deaf people. Prelingually deaf people acquire the ability to speak with great difficulty and then their speech is generally considered to be quite abnormal. Postlingually deaf people experience deterioration of voice F_0 and amplitude control shortly after the onset of hearing loss (Binnie *et al.*, 1982), while control of articulatory dynamics declines more slowly.

There have been several different experimental approaches towards the study of the role of auditory feedback on voice control; the Lombard effect, side-tone amplification (Lane and Tranel, 1971), noise masking (Elliott and Niemoeller, 1970; Ward and Burns, 1978) and perturbation (Sapir *et al.*, 1983). Recent studies have demonstrated that vocalizing subjects compensate for a perturbation in voice pitch feedback by changing their voice F_0 in the opposite direction to the change in feedback. The compensatory nature of the responses suggest that they are an attempt to correct for an error between voice pitch feedback and the note the subject was attempting to produce. Such responses have been observed during sustained vowel sounds, glissandos, speech, and singing (Bauer, 2004; Bauer and Larson, 2003; Burnett *et al.*, 1998; Burnett and Larson, 2002; Donath *et al.*, 2002; Hain *et al.*, 2000; Hain *et al.*, 2001; Jones and Munhall, 2000, 2002; Kawahara and Williams, 1996; Kiran and Larson, 2001; Larson, 1998; Larson *et al.*, 2001; Larson *et al.*, 2000; Natke *et al.*, 2003; Natke and Kalveram, 2001; Sivasankar *et al.*, 2005; Xu *et al.*, 2004). In approximately 15% of the trials, subjects produced responses that changed in the same direction as the stimulus, e.g., an upward change in voice F_0 in response to an upward perturbation in voice pitch feedback. These have been termed "following" responses.

We described and modeled the system responsible for generating responses to pitch-shifted voice feedback as a negative feedback control system (Hain *et al.*, 2000). Elements of this system must have an internal referent of desired voice F_0, access to feedback pitch and the ability to correct for errors between the pitch of the feedback signal and the referent. It has also been demonstrated that perturbation of voice loudness feedback results in compensatory responses in voice amplitude (Bauer *et al.*, 2006; Heinks-Maldonado and Houde, 2005). These voice amplitude responses have the same approximate latency as F_0 responses to pitch-shifted feedback and are compensatory in nature. We have also presented a model and described the system responsible for the

Schematic depiction of stimulus conditions for the three experiments
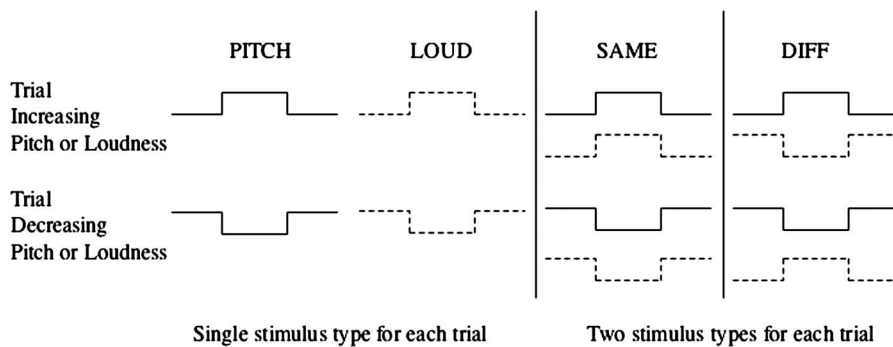
FIG. 1. Schematic representation of the stimuli used in the four experimental conditions. Square brackets depict the direction and relative timing of pitch-shifted (solid lines) or loudness-shifted (dashed lines) feedback.

loudness-shift responses as a negative feedback control system with properties similar to those of the voice F_0 control circuitry.

Although it appears that both pitch and loudness perturbations elicit similar responses, it is unknown whether the two responses are independent of each other, and thus represent different mechanisms, or whether they are part of the same responding system. We tested the hypothesis that the two responses are independent of each other by presenting to subjects vocalizing a vowel sound, simultaneous changes in pitch and loudness feedback, where both of the stimuli changed in the same direction or in the opposite direction. We predicted that if there are two independent mechanisms, then a stimulus composed of both pitch- and loudness-shifted voice feedback should elicit two independent responses, an F_0 response that changed in the opposite direction to the pitch-shift stimulus and an amplitude response that changed in the opposite direction to the loudness-shift stimulus. As control studies, subjects were also tested with pitch- and loudness-shift stimuli alone.

Results of the experiments revealed the presence of both F_0 and amplitude responses to simultaneous pitch- and loudness-shifted voice feedback that generally changed in opposite directions to the respective stimuli. Thus, the results suggest that there are two separate systems controlling F_0 and amplitude responses to pitch- and loudness-shifted stimuli and that the mechanisms underlying voice F_0 and amplitude control based on pitch and loudness feedback are independent. Additional observations demonstrate that with some stimuli there are interactions between the two response mechanisms.

II. METHODS

A. Subjects

Twenty-four undergraduate students at Northwestern University (2 male, 22 female, ages 18–22 years) served as subjects. Similar studies conducted in our laboratory have not revealed significant differences in responses as a function of sex, and hence we did not attempt to recruit equal numbers of male and female subjects. All subjects passed a hearing screening, and none reported a history of neurological or communication disorders. All subjects signed informed consent approved by the Northwestern Institutional Review Board. After preliminary data analysis, data from one subject were excluded because of an incomplete data set. Final results are based on data from 23 subjects (1 male, 22 female).

B. Apparatus

Subjects were seated in a sound-treated room and wore Sennheiser headphones with attached microphone (Model No. HMD 280). The vocal signal from the microphone was amplified with a Mackie mixer (Model No. 1202), processed for pitch and loudness shifting with an Eventide Eclipse Harmonizer, mixed with 40 dB SPL pink masking noise with a Mackie mixer (Model No. 1202-VLZ), further amplified with a Crown D75 amplifier and HP dB attenuators, and then sent back to the headphones. The harmonizer was controlled with MIDI software (Max/MSP v4.5 by Cycling '74) from a laboratory computer. Acoustic calibrations were made with a B&K 2250 sound level meter and model 4100 in-ear microphones. There was a gain of 10 dB SPL between the subject's voice amplitude, measured 2.5 cm from the mouth, and the feedback loudness measured at the input to the ear canal. The delay from MIDI input to the harmonizer to the shift in pitch was 14 ms, whereas there was no delay for the loudness shift stimulus. We considered the 14 ms difference in processing speed of the two feedback sources to be too small to affect the results. Preliminary testing revealed no perceptible difference in the timing of the pitch- and loudness-shifted feedback. The voice output signal, feedback and control pulses were digitized at 10 kHz, low-pass filtered at 5 kHz and recorded on a laboratory computer utilizing Chart software (ADInstruments, Colorado Springs, CO). Data were analyzed using event-related averaging techniques in Igor Pro (Wavemetrics, Inc., Lake Oswego, OR). Subjects monitored their voice loudness from a Dorrough Loudness Monitor placed 0.5 m in front of them. This monitor provided the subjects with visual feedback of their voice amplitude and helped them to maintain a relatively constant level throughout the testing.

C. Procedures

Subjects produced /u/ vowel sounds into a microphone while hearing their voice modulated in pitch or loudness over headphones in near real-time (Bauer and Larson, 2003; Bauer et al., 2006). Each vocalization was ≈5 s, in duration and produced at a level of ≈70 dB. Voice feedback loudness was amplified to ≈80 dB SPL. During each trial (vocaliza-
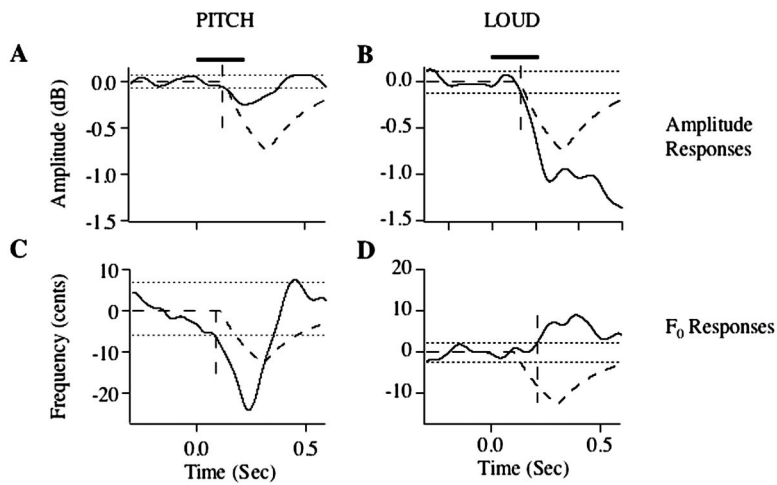
FIG. 2. Averaged F_0 and amplitude contours for the PITCH and LOUD conditions. Top row, responses to upward directed stimuli with voice amplitude responses. Bottom row, responses to upward directed stimuli with voice F_0 responses. Horizontal dashed lines represent ±2 SDs of the prestimulus mean. Stimulus onsets were at time 0.0, indicated by horizontal bar above the plots. Short vertical dashed lines indicate onset of response. Curved dashed traces are simulations from the model.

tion), five stimuli were presented at randomized intervals of 0.7–1.0 s. For each condition consisting of 8 trials, a total of 20 increasing and 20 decreasing stimuli were presented to the subjects. In the PITCH condition, randomized ±50 cent pitch-shift stimuli were presented, and in the LOUD condition, randomized ±3 dB loudness-shifted stimuli were presented. For any given set of 8 trials, the stimulus type (pitch or loudness) was held constant. In the SAME condition, simultaneous pitch and loudness-shifted stimuli were presented, and they both either increased or decreased. In the DIFF condition, simultaneous pitch- and loudness-shifted stimuli were presented, and they changed in different directions. For example, a +50 cent pitch shift was combined with a −3 dB loudness shift. Figure 1 diagrammatically illustrates the stimuli for the three conditions.

From the digitized signals, one wave representing the F_0 contour and one representing the voice amplitude contour were generated (Bauer and Larson, 2003; Bauer *et al.*, 2006). Event-related averages were generated for each subject for each experimental condition by time aligning the voice F_0 and amplitude contours with the pitch-shift or loudness-shift stimulus onset (TTL control pulse). For each type of stimulus, event-related averages were calculated for both F_0 and amplitude contours. Each averaged response consisted of a minimum of 15 trials, with a 200 ms prestimulus baseline and a 500 ms poststimulus response window. Valid responses were identified according to the following criteria: a deviation in the averaged trace (F_0 or amplitude) with a magnitude >2 standard deviations (SDs) of the pre-stimulus baseline for a minimum duration of 50 ms, and a latency ⩾60 ms after stimulus onset (see Fig. 2). Response latency was defined as the time point where the averaged trace crossed the 2 SD line following the stimulus. Response magnitude was measured as the greatest magnitude of the averaged F_0 or amplitude trace from the baseline mean. In the SAME and DIFF conditions, since each stimulus consisted of a combined pitch and loudness component, both voice amplitude and F_0 trajectories were averaged for each stimulus combination. Statistical analyses were done only on compensating responses because "following" responses may represent errors in responding, as explained in the following. Response latencies and magnitudes were submitted to statistical testing

with one-way ANOVAs (Data Desk; Data Description). Excessive numbers (>8) of missing responses in the −3 dB conditions precluded a repeated-measures design. Response directions (compensating or "following") were tabulated by condition and stimulus direction.

III. RESULTS

Subjects responded in both PITCH and LOUD conditions with changes in F_0 and voice amplitude. Although most responses were compensatory, many "following" responses were also produced, primarily with cross-dimensional stimuli. Table I displays the number of F_0 and amplitude "compensatory," "following," and nonresponses (NR) in the PITCH and LOUD conditions. Most subjects produced a change in F_0 or amplitude in response to either a perturbation in pitch or loudness feedback, however, there were differences in the types and number of responses across conditions. In the PITCH condition, 87% of F_0 responses were compensatory, whereas 50% of amplitude responses were compensatory. In the LOUD condition, 52% of F_0 responses were compensatory while 67% of amplitude responses were compensatory.

Figure 2 displays illustrative examples of averaged F_0 and voice amplitude responses for both PITCH and LOUD conditions from two different subjects. As can be seen, all responses are in the compensatory direction except for Fig. 2(D), which is a "following" response. The cross-dimensional responses, i.e., amplitude response to a pitch-shift stimulus [Fig. 2(A)] and F_0 response to a loudness-shift

TABLE I. Numbers of compensatory (COMP), "following" (FOL) and non-responses (NR) for Pitch and Loud conditions and stimulus dimension type[a].

	PITCH		LOUD		
Response type	F_0	Amplitude	F_0	Amplitude	Total
COMP	40 (87%)	23 (50%)	24 (52%)	31 (60%)	118 (64%)
FOL	5 (11%)	15 (33%)	17 (37%)	9 (20%)	46 (25%)
NR	1 (2%)	8 (17%)	5 (11%)	6 (13%)	20 (11%)
Total	46	46	46	46	184

[a]Chi square=13.21, df=2, $p < 0.002$.

TABLE II. Counts of types of voice amplitude responses across conditions. "Following" (FOL), nonresponses (NR), and compensatory (COMP) responses are tabulated by simultaneous pitch- and loudness-shifted feedback with the stimuli changing in the same (SAME) or different (DIFF) directions. Responses are further organized with each specific stimulus combination of either + of −50 cent and + or −3 dB stimuli.[a]

| | SAME | | DIFF | | |
	−50 c −3 dB	+50 c +3 dB	−50 c +3dB	+50c −3dB	Total
COMP	12 (52%)	20 (87%)	20 (87%)	12 (52%)	64 (70%)
FOL	8 (35%)	0	2 (9%)	7 (30%)	17 (18%)
NR	3 (13%)	3 (13%)	1 (4%)	4 (17%)	11 (12%)
Total	23	23	23	23	92

[a]Chi-square=16.26, df=6, $p=0.0124$.

TABLE III. Counts of types of F_0 responses across conditions. "Following" (FOL), nonresponses (NR), and compensatory (COMP) responses are tabulated by simultaneous pitch- and loudness-shifted feedback with the stimuli changing in the same (SAME) or different (DIFF) directions. Responses are further organized with each specific stimulus combination of either + of −50 cent and + or −3 dB stimuli.[a]

| | SAME | | DIFF | | |
	−50c −3dB	+50c +3dB	+50c −3dB	−50c +3dB	Total
COMP	20 (87%)	20 (87%)	16 (70%)	13 (57%)	69 (75%)
FOL	1 (4%)	2 (9%)	6 (26%)	8 (35%)	17 (18%)
NR	2 (9%)	1 (4%)	1 (4%)	2 (9%)	6 (7%)
Total	23	23	23	23	92

[a]Chi-square=10.39, df=6, $p=0.1093$.

stimulus [Fig. 2(D)] are smaller than the within-dimensional responses (also see Fig. 5). There were also instances where a double or triple response was observed. In such cases, we measured the first compensatory response for statistical analysis. If compensatory responses were not made, we measured the first "following" response. Responses to downward pitch- and loudness-shifted stimuli were similar to those illustrated in figure 2 but the directions were reversed.

Tables II and III provide counts of compensating, "following," and nonresponses for amplitude (Table II) and F_0 (Table III) responses for the SAME and DIFF conditions. For these conditions, as all stimuli consisted of both a pitch and loudness component, we defined compensating and "following" F_0 and amplitude responses with respect to stimuli that were in the same acoustical dimension. For example, an upward voice amplitude response to a stimulus combination of +50 cent −3 dB, would be classified as a compensating response because its direction is opposite to the downward 3 dB component of the stimulus. For the amplitude responses, the types of responses differed significantly across the four stimulus combinations comprising the SAME and DIFF conditions. There was a disproportionately large number of amplitude "following" responses in the stimulus combinations that included a −3 dB stimulus (35% for SAME and 30% for DIFF). Otherwise, response types were rather

evenly distributed across the stimulus combinations. For the F_0 responses, although not significantly different, there were more "following" responses in the DIFF (30%) than in the SAME (7 %) condition. Altogether, 18% of responses in the SAME and DIFF conditions were of the "following" type, compared with 25% for the PITCH and LOUD conditions. Compensating responses constituted 72% of the SAME and DIFF responses.

Figure 3 illustrates representative amplitude (top) and F_0 responses (bottom) to simultaneous stimuli changing in the SAME direction. Upward stimuli on the left and downward stimuli on the right led to opposing directed responses in each case. The amplitude response to a stimulus combination including a −3 dB stimulus (B) was much smaller than that to a combination including a +3 dB stimulus. The F_0 responses are roughly of the same magnitude for each stimulus. Response latencies in all of these cases are approximately 100 ms.

Figure 4 shows representative responses in the DIFF condition. On the left, the stimulus (+50 cent −3 dB) led to an upward amplitude response (A), which opposed the −3 dB stimulus, and a downward F_0 response (C) that opposed the +50 cent stimulus. On the right, with the −50 cent +3 dB stimulus, again both the amplitude (B) and F_0 responses (D) oppose the direction of the stimulus that was in
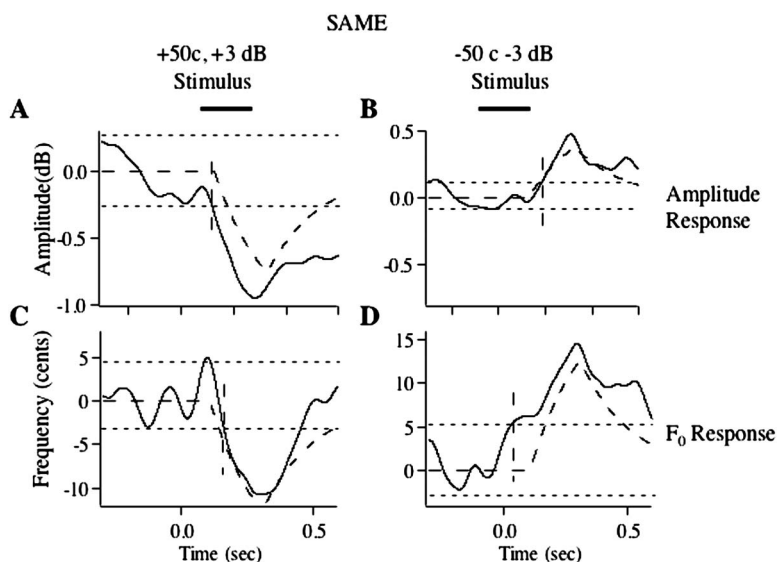
FIG. 3. Averaged F_0 and amplitude contours for the SAME condition. Top row shows contours of amplitude and bottom row of F_0 responses. Left column shows responses to stimuli composed of +50 cent and +3 dB stimulation. Right column shows responses to stimuli composed of −50 cent and −3 dB stimulation. Curved dashed lines are simulations from the model.
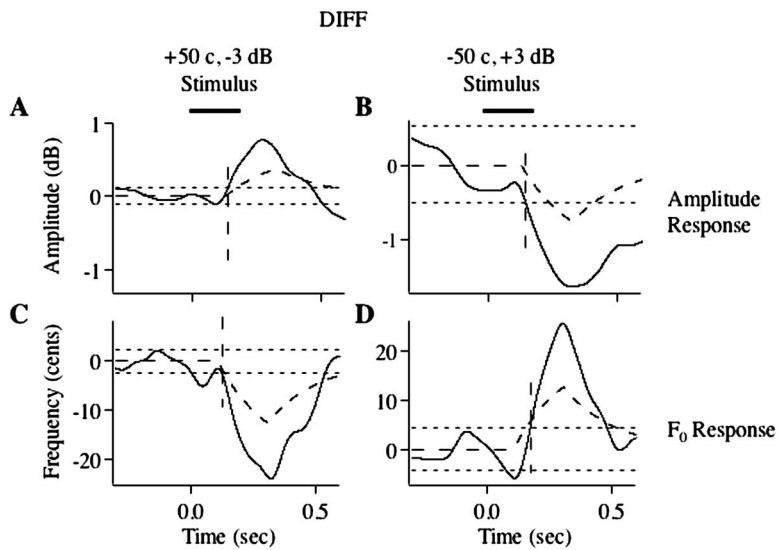
FIG. 4. Averaged F_0 and amplitude contours for the DIFF condition. Top row shows contours of amplitude and bottom row of F_0 responses. Left column shows responses to stimuli composed of +50 cent and −3 dB stimulation. Right column shows responses to stimuli composed of −50 cent and +3 dB stimulation. Curved dashed lines are simulations from the model.

the same acoustical dimension. There were also occasions when a small apparent "following" response occurred prior to a much larger compensating response [Figs. 4(C) and 4(D)]. It is unclear what these responses represent; whether they just represent 'noise' in the wave forms or a bona fide response. It is noteworthy that the amplitude response to the +50 cent −3 dB stimulus is much smaller than the response to the −50 cent +3 dB stimulus. This difference is similar to the comparison of amplitude responses in Fig. 3.

Figure 5 displays the overall mean response magnitudes for both voice amplitude and F_0 responses across all four experimental conditions, PITCH, LOUD, SAME, DIFF, and the two types of stimuli within each condition. Statistical testing across all four conditions revealed an effect for voice F_0 magnitude ($F = 8.77$, df = 3, 128, $p < 0.0001$). Posthoc testing with a Bonferroni correction indicated the mean response magnitude for the PITCH condition (18 cents) was signifi-

cantly larger than for the SAME (13 cents; $p < 0.007$) and the LOUD condition (10 cents; $p < 0.0001$) but not different from the DIFF condition (16 cents). There was no significant effect for pitch-shift direction. For magnitude of voice amplitude responses, there was no significant difference across conditions, however, the downward loudness-shift stimulus (−3 dB) elicited significantly smaller responses (mean 0.54 ± 0.23 dB) than the upward stimuli (mean 0.91 ± 0.35 dB; $F = 30.11$, df = 1, 93, $p < 0.0001$). Although it would be desirable to compare F_0 response magnitudes with amplitude magnitudes, their differing dimensions preclude this.

For voice amplitude magnitudes across all eight conditions, there was an overall effect ($F = 6.16$, df = 7,110, $p < 0.0001$). Posthoc testing with Bonferroni corrections showed that loudness-shift stimuli of −3 dB led to smaller responses than all of the following stimulus combinations:
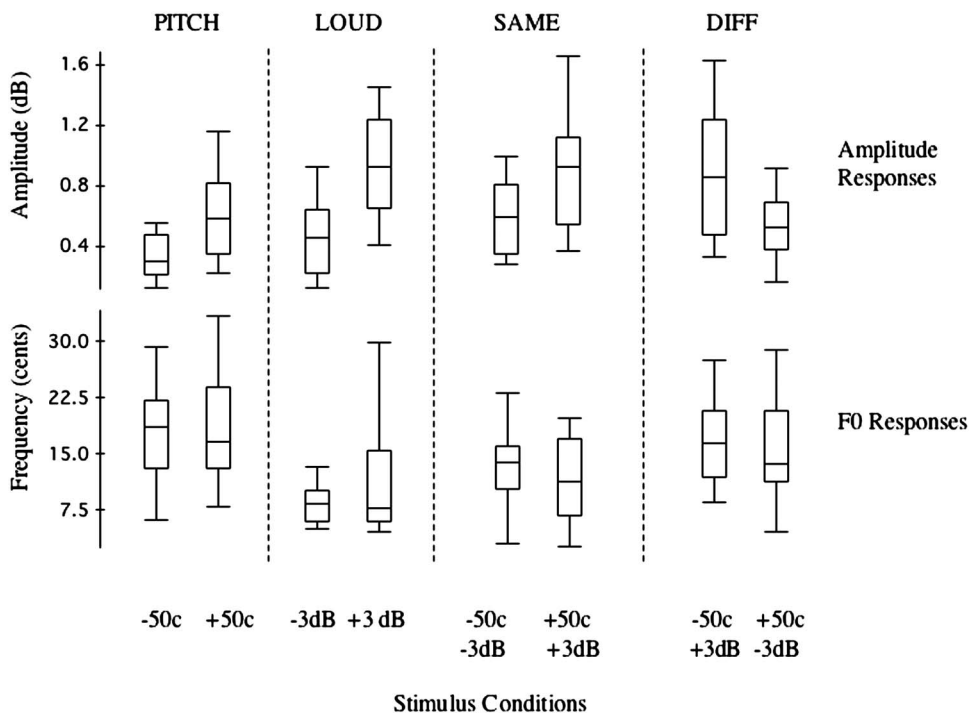
FIG. 5. Boxplots of response magnitudes for all experimental conditions. Top row shows amplitude responses and bottom row F_0 responses. Box definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge − low hinge), and low hinge −1.5 (high hinge − low hinge) (data desk; data description).
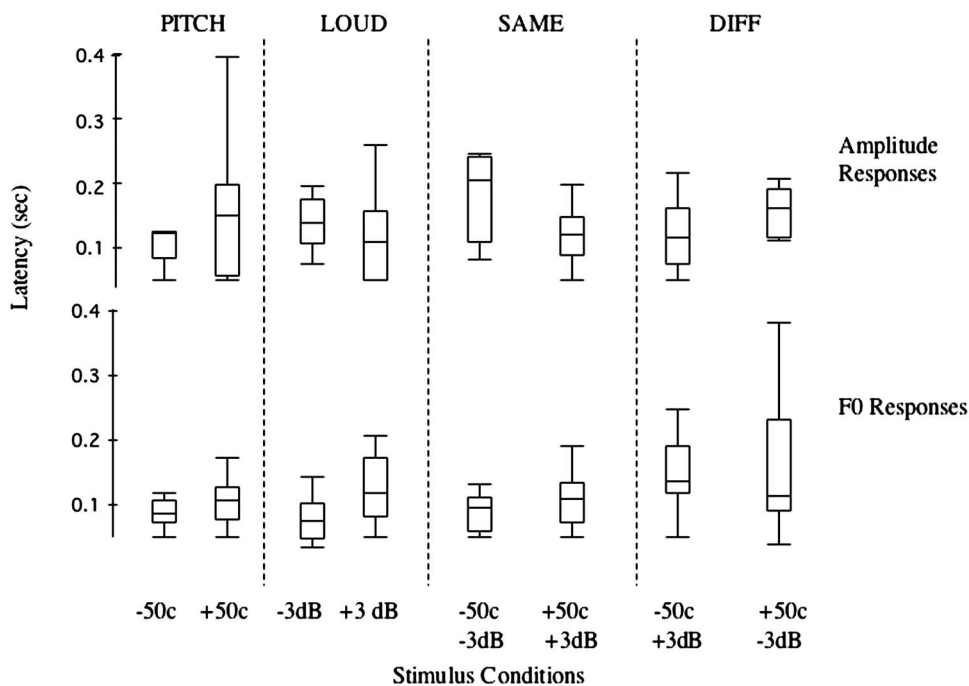
FIG. 6. Boxplots of response latencies for all experimental conditions. Top row shows amplitude responses and bottom row F_0 responses.

−50 cent +3 dB ($p < 0.04$), +3 dB ($p < 0.004$), and +50 cent +3 dB ($p < 0.04$). Stimuli of −50 cents also produced smaller responses than the following stimulus combinations: −50 cent +3 dB ($p < 0.02$), +3 dB ($p < 0.004$), and +50 cent +3 dB ($p < 0.02$). Finally, the stimulus combination of +50 cent −3 dB produced a smaller response than +3 dB stimuli ($p < 0.008$).

Figure 6 displays the mean latencies for responses across all four experimental conditions and stimuli. Statistical testing revealed an effect of the testing condition for F_0 responses ($F = 4.13$, df = 3,128, df < 0.008). Posthoc testing with a Bonferroni correction showed that the latencies of the DIFF condition (mean 152±82 ms) were significantly longer than those for the PITCH (mean 106±64 ms; $p < 0.02$) and the SAME conditions (mean 104±42 ms; $p < 0.02$). There were no significant differences for F_0 response latencies as a function of direction of the pitch-shift stimulus.

There was no significant difference in voice amplitude response latencies as a function of condition, however, the downward loudness-shift stimulus produced significantly longer latencies (mean 163±49 ms) than did the upward stimulus (mean 125±63 ms; $F = 9.68$, df = 1,93, $p < 0.003$). Mean latencies of F_0 responses (mean 128±81 ms) were significantly shorter than the amplitude latencies (mean 155±91 ms; $F = 8.7$, df = 1,328, $p < 0.005$).

Statistical testing of response latencies across all conditions shown in Fig. 6 showed a main effect for F_0 latencies ($F = 2.56$, df = 7,124, $p < 0.02$), but posthoc testing with a Bonferroni correction did not reveal significant differences between any two stimulus combinations. There was no significant difference for voice amplitude response latencies as a function of the stimulus combinations shown in Fig. 6.

We simulated these responses, as shown in Figs. 2–4 (curved dashed lines), using the negative feedback model shown in Fig. 7. This model, implemented with the SIM-ULINK toolbox of MATLAB (Mathworks, Nantick, MA) is a combination of two previously reported feedback models, one of stabilization of F_0 (Hain et al., 2000), and the other of stabilization of loudness (Bauer et al., 2006). Desired loudness and F_0 are compared to perceived loudness and F_0, and an error signal is computed through subtraction. Next, the error signal is "demultiplexed," meaning it is converted from a tonotopic representation to scalers representing F_0 and loudness error. Then, in the "pitch error feedback" and "loudness error feedback" sections, error is delayed and then used to adjust voice drive. Thus there are two negative feedback models, linked together as necessitated by known physiology, by multiplexers and demultiplexers for F_0 and loudness. The goal of the aggregate model was to establish feasibility and to provide a quantitative hypothesis for future work. We portray simulations obtained from a set of generic parameters optimized to fit the data shown in figures 2–4.

The previous individual models of F_0 and loudness were combined by explicitly modeling the computation of F_0 and loudness from auditory input (Bauer et al., 2006; Hain et al., 2000). In previous models we simply assumed that these signals were available. In the present model we consider F_0 and loudness as being "multiplexed" together in the cochlea. We made no assumption concerning whether the encoding is via a tonotopic representation, phase-locking, or a combination of the two. We then "demultiplex" the central auditory signal into F_0 and its loudness in a process that takes about 100 ms ("demux delay" in Fig. 7). In physiological terms, this would mean that the representation of F_0 and the loudness of F_0 are not separated from each other until a lengthy decoding step occurs, presumably in auditory cortex. A "matrix gain" cross-coupling matrix was included to allow for the possibility of interactions between F_0 and loudness, but it was configured without cross coupling as this simple configuration fit our data reasonably well.

Although we used a lumped demux delay of 100 ms, the true delay associated with the conversion of the central rep-

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Larson et al.: Perturbations of voice pitch and loudness   2867

Pitch Error Feedback
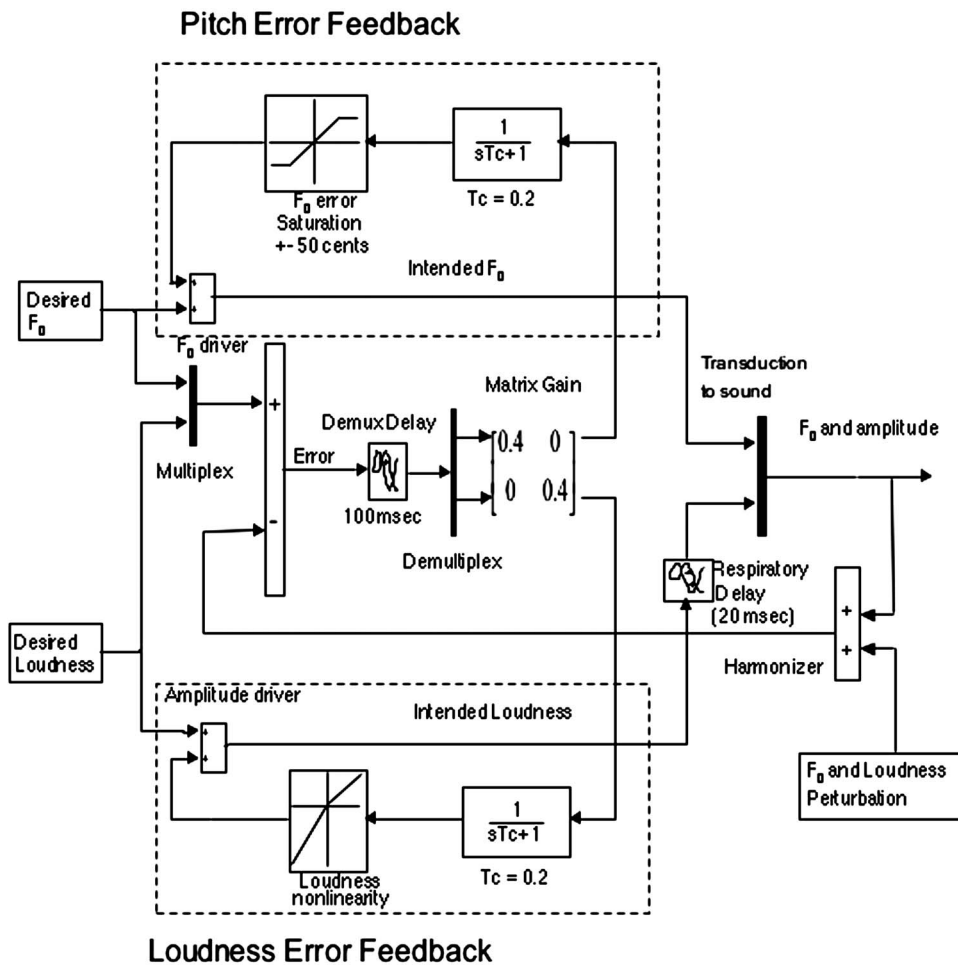
FIG. 7. Model of negative feedback system for control of voice F_0 and amplitude responses to pitch and loudness-shifted voice feedback. See the text for further details.

Loudness Error Feedback

resentation into the scalars of F_0 and loudness error must be less, as the 100 ms includes time to get the signal to the cortex as well as sending it back to the brainstem. The 100 ms delay of the demux delay could be made shorter without affecting the simulations by shifting some of this delay to these pathways, but as our simulations would be unaffected by the shifts, for simplicity we have lumped them together. Physiological data suggests that there is likely a delay of about 20 ms between auditory input and auditory cortex (Howard *et al.*, 1996; Steinschneider *et al.*, 1999) and that the latency between cortical stimulation and activation of laryngeal and respiratory muscles is on the order of 13 ms (Gandevia and Rothwell, 1987; Ludlow and Lou, 1996). Thus, the anatomical demux delay should be about 66 ms.

We also attempted to model the longer latency observed for perturbations of loudness. Because the loudness signal is the loudness of F_0, only after F_0 has been determined can the loudness of F_0 be computed. This is most easily seen by an example—if overall loudness were used for negative feedback, then speech in noise would be softer than speech in quiet, because the brain would not be able to distinguish loudness of noise from speech. However, we know that the opposite is true from the Lombard effect literature (Lane and Tranel, 1971). It follows that because the relevant signal is the loudness of F_0, not loudness in general, and that the loudness feedback step must await the F_0 determination step. We implemented the additional loudness delay as "respiratory delay" in the loudness output pathway.

In order to simulate our observations that responses to upwards perturbations of loudness were stronger than downward perturbations, we added a nonlinearity to the loudness portion of the model, with a higher gain for upward perturbations. This nonlinearity has a slope of 0.5 for downward perturbations and 1.0 for upward perturbations, to create the roughly 2:1 gain asymmetry observed experimentally.

A model sensitivity analysis was performed to ascertain whether there was an optimal set of gain and time constant parameters as well as whether or not model performance was critically dependent on one or the other. The analysis revealed that best performance in terms of the model in a typical subset of the experimental data, was found with the gain and time constant parameters set to 0.4, and 0.2, respectively. This was the case for *both* the F_0 and loudness subsystems, and accounted for about 70% and 50% of the variance, respectively. Very poor fits occurred for the combination of large gains (e.g., 2) and short time constants (e.g., 0.2–0.6). The similarity between optimal fits for the F_0 and loudness stabilization subsystems suggests that they share neural circuitry.

IV. DISCUSSION

In this study, changes in voice F_0 and amplitude were measured in response to either pitch- or loudness-shifted voice feedback when the stimuli were presented alone or when both pitch- and loudness-shifted feedback were pre-

sented simultaneously. Other studies have shown that with dynamically changing pure tones, there may be interactions between the pitch and loudness stimuli, and in some cases these are associated with perceptual illusions (McBeath and Neuhoff, 2002; Moore and Sek, 1998; Neuhoff *et al.*, 1999). To our knowledge, no other studies have examined these relationships with respect to voice feedback. The present study was conducted to answer the question of whether or not the voice F_0 and amplitude responses to pitch- and loudness-shifted stimuli represent the same, or two independent mechanisms. By first testing the subjects with either pitch-or loudness-shifted stimuli (PITCH and LOUD), it was demonstrated that the basic properties of the F_0 and amplitude responses to stimuli in the same acoustical dimensions are very similar in form and are optimally simulated by identical mathematical constructs. This suggests that the two mechanisms share neural circuitry. Nevertheless, our results from experiments where both stimuli were present simultaneously, show that the two systems can respond largely independently of each other, albeit with small nonlinear interactions and minor cross coupling. Thus, overall it appears that the two stabilization mechanisms are predominantly independent but share circuitry and interact to a minor extent.

A. Theoretical basis for interactions between pitch and loudness processing

The responses to pitch- or loudness-shift stimuli result from a three-step process involving conversion of the stimuli into neural firing patterns, central processing and motor output. Given the fact that cross-dimensional responses were produced, a mixing together between loudness and pitch processing must occur at one of these three steps. The "mixing together" might consist of a linear combination of the two signals, which we call "cross coupling," or a nonlinear influence of one signal on the other, which we call "nonlinear interactions." In general, mixing together between two sensory modalities causes confusion between sensory streams and is undesirable. Thus, the work reported here is essentially looking for design flaws in the auditory system.

To determine where the mixing might occur requires a more detailed consideration of how the brain processes auditory signals and in particular, how it detects changes in pitch or loudness. Sound is not transmitted from the cochlea to auditory cortex as separate streams of "F_0" and "loudness," but rather it is tonotopically encoded in the cochlea. In other words, F_0 and loudness are multiplexed together. There is also processing of F_0 and loudness in the brainstem. As the signal most relevant to F_0, "phase locking," is not found in auditory cortex above about 100 Hz(Palmer, 1995), cortical processing probably uses tonotopic signals. If we accept that F_0 is necessarily computed in cortex, the loudness of F_0 cannot be computed until the cortical step has finished, and thus, although the brainstem could potentially compute overall loudness as well as the loudness of spectral components, it is highly unlikely to be the location for computation of the loudness of F_0. It is generally agreed that perception of sound depends upon the representation of individual tonotopic components being "reassembled" at a later stage in the auditory system (Moore, 1995). A lengthy reassembly (de-

multiplexing) step is plausible if one considers that converting from a tonotopic array representing cochlear input into separate signals encoding F_0 and loudness of F_0 intrinsically requires first F_0 identification, followed by computation of the loudness of F_0. F_0 identification is necessary because the cochlea cannot determine, without the help of the cortex, which tonotopic signal of an incoming frequency spectrum is the fundamental frequency.

Although a simple method of finding the largest component of the spectrum would work perfectly well for a pure sinusoid, there is overwhelming evidence that this is not the usual method used by the auditory system. It is well known that F_0 can be "heard" even if it is physically absent, as long as there are associated higher harmonics (Schouten, 1940). Also, two fundamentals can be separated out from other, potentially louder pitches, as for example, when we listen to a chord being played by a group of instruments (Houtsma, 1995). These observations teach us that simple algorithms such as finding the loudest component of a spectrum cannot be those used to derive F_0, and also that the cortex must participate in the process.

Previous central models of complex pitch detection, such as the "optimum processor model" of Goldstein, postulated that F_0 is identified via template matching (Goldstein, 1973). Template matching, being a high-level process, would have to be implemented in auditory cortex as well as be relatively slow. Once F_0 is determined, then the loudness of F_0 might be computed by selecting the tonotopic input of F_0, which would be relatively fast. Alternatively, considering Schouten's residue theory (Schouten, 1940), F_0 determination might emerge from the template matching process itself. Either method would seem likely to be a much quicker process than template matching alone. Thus, it seems likely that F_0 and loudness processing relevant to feedback control should share considerable common neural circuitry, up through a lengthy step involving F_0 identification and demultiplexing.

Nevertheless, our data provides compelling evidence that the F_0 and loudness systems are predominantly independent. The fact that the responses were able to respond appropriately to within-dimensional stimulation and change in a direction opposite to the within-dimensional stimulus, and opposite in direction to each other, provides strong evidence that the mechanisms controlling the F_0 and amplitude responses *can* function independently. Depending on the stimulus combinations, both responses can change in the same direction or in opposite directions.

B. Linear interaction between loudness and F_0 processing

As mentioned previously, using engineering formalism, the mixing together between F_0 and loudness—a sensory or motor confusion—can be thought of as two types of basic processes—linear combination (dimensional cross coupling) and nonlinear interactions. Dimensional cross coupling is portrayed in our model as a 2×2 matrix located after error has been computed and the tonotopic representation of F_0 and loudness have been demultiplexed. A matrix with large

elements on the diagonals compared to the off-diagonals, would correspond to a system without cross coupling. A matrix with large off-diagonal elements, corresponds to a system with cross coupling. Our data suggests that there is a small off-diagonal element between F_0 and voice amplitude (element 1,2 of the gain matrix), and that there is no off-diagonal coupling between loudness and F_0 (element 2,1). Further, our simulations with the model configured with no cross coupling or nonlinear interactions accounted for most of the variance in the experimental data. Thus, our results suggest that representation of F_0 and a loudness stabilization system as predominantly independent is feasible.

Logically, given that auditory input is transmitted to the cortex using tonotopic encoding which keeps F_0 and amplitude combined, mixing could occur at any point—input, central processing, or motor output. Previous literature supports the suggestion that there may be a mixing together, which is appreciated at the perception stage (McBeath and Neuhoff, 2002; Moore and Sek, 1998; Neuhoff *et al.*, 1999). Perception includes both input and central processing. Thus, in the context of our model, these observations support the idea that a part or all of the mixing together does not occur in the output circuitry.

C. Nonlinear interactions between F_0 and loudness processing

The other potential type of mixing together between modalities is nonlinear interactions. Examples of nonlinear interactions might be an influence of one sensory modality on the timing of another, or a multiplicative type action of one sensory modality on the other. Like cross coupling, nonlinearity is generally avoided in system design. However, there was evidence for nonlinear interactions in our data.

The primary finding of the amplitude responses was that all stimuli that included a −3 dB component were smaller and slower than those produced to other stimulus combinations. The small responses to the −3 dB stimulus, represented by the "loudness nonlinearity" in our model, may be related to the observation that for equal differences in an increase or a decrease in the intensity of sounds, subjects perceive an increase to be greater than a decrease (Neuhoff *et al.*, 2002). It is also known that a decrease in stimulus intensity leads to longer reaction times (Jaskowski *et al.*, 1994; Seitz and Rakerd, 1997), which occurred with the −3 dB stimuli. Along this line, it is instructive that the amplitude responses were also much smaller for a −50 cent pitch-shift stimulus than a +50 cent stimulus. The most parsimonious explanation is that the same nonlinearity that is responsible for smaller responses to decreases in intensity, also acts on the cross-coupled response to F_0.

There was additional evidence for nonlinear interactions between the response mechanisms. When a loudness-shift stimulus was combined with a pitch-shift stimulus in the same direction (SAME), F_0 response magnitudes were smaller than those made to a pitch-shift stimulus alone (PITCH). In this condition there was a change in both pitch and loudness, and they changed in the same direction. Yet, despite the fact that both pitch and loudness were dropping,

the F_0 responses were much smaller than those to a downward pitch-shift stimulus alone. In other words, there is a nonlinear interaction between the direction of change of inputs and the size of the response. It might be argued that with both a decrease in pitch and loudness, the overall perceptual impression would have been that of a decrease in vocal effort, such as a drop in voice amplitude and frequency that occurs in a nonfocused syllable (Alain, 1993; Cooper *et al.*, 1985; Eady and Cooper, 1986), and that such a "natural" event might be intended and thus should not trigger a response.

Observations from the DIFF conditions also document nonlinear interaction between the two mechanisms. In the DIFF condition, F_0 responses were somewhat smaller than in the PITCH condition, but much slower than in the other conditions.

The observations of "following" responses also suggest nonlinear interactions between the two response mechanisms. We have previously suggested that "following" responses occur due to a self-selected tracking mode where the subjects consider auditory feedback to be externally generated (as it is), rather than self-generated (Hain *et al.*, 2000). Most "following" voice amplitude responses occurred whenever the stimuli included the −3 dB component, and most "following" F_0 responses occurred when a pitch- and loudness-shift stimulus were presented in conflict with each other (DIFF). These observations suggest that the −3 dB stimulus may have been near threshold of detection of loudness, and the responses were almost random. "Following" F_0 responses were most prevalent when a pitch and loudness shift stimuli were combined in opposite directions. This stimulus combination may have made stimulus identification difficult; the responses were slower, smaller and more easily confused with "noise" in the system. This behavior may occur when the stimulus sounds unlike the subject's own voice, which could have been caused by some of the stimulus combinations in the present study. Regardless of the cause of "following" responses, the fact that they occurred under different conditions for amplitude and F_0 dimensions, suggests that the control circuitry for these two response mechanisms must not be completely combined.

D. Evidence relating to separation of F_0 and loudness control circuitry

Other observations from this study suggest there are differences in these two response mechanisms, not having to do with interactions that would support the idea that there is at least partial separation of the control circuitry. The finding that F_0 latencies (mean 116 ms) were significantly shorter than amplitude latencies (mean 150 ms) is one such example. In the calculation of these latencies, the difference in timing of the actual pitch- or loudness-shifted feedback that was caused by the harmonizer was taken into account. The latency differences cannot be entirely due to differences in transmission delays from the cortex to the respiratory or laryngeal muscles, since these delays are both about 13 ms (Gandevia and Rothwell, 1987; Ludlow and Lou, 1996). The latency differences may be due to differences in the percep-

tual saliency of the two stimuli. Both the ±3 dB and ±50 cent stimuli were readily apparent to the investigators, but no perceptual testing was done to determine if one or the other stimulus was more or less apparent to the subjects in the study. It is known that increases in stimulus intensity or perceptual salience can reduce the latency of responses to acoustical stimuli (Jaskowski *et al.*, 1994; Klug *et al.*, 2000), and if the subjects perceived the ± 50 cent stimuli to be more salient than the ±3 dB stimuli, this could account for the shorter latencies of the pitch-shift stimuli.

These latency differences could also be due to the motoric mechanisms controlling the responses. It is likely that the F_0 responses were controlled by laryngeal muscle contractions; the cricothyroid muscles are primarily responsible for regulating voice F_0 (Titze, 1994). Moreover EMG studies have shown the cricothyroid muscle to respond to pitch shift stimuli (Kawahara *et al.*, 1993). Voice amplitude responses could be due to respiratory muscle contractions and/or laryngeal muscle contractions. If they are due to contractions of muscles such as the intercostals, there may be a significant time delay between muscle contraction, thoracic volume reduction and compression of the broncho-tracheo air column. On the other hand, if the amplitude responses are due to changes in laryngeal impedance or stiffness, which is controlled primarily by intrinsic laryngeal muscles (Titze, 1994), then there may be no differences in response latency based on muscle contractions. Additional studies are necessary to determine which muscles are activated by pitch- and loudness-shift stimuli. It is also possible that the latency differences could be due to cortical processing delays, which are largely unknown at this time. We implemented the longer latency for loudness shifts in our model of the response as a separate delay (respiratory delay) following F_0 extraction.

E. Implications for our understanding of complex vocal control

Aside from providing information about mechanisms involved in generating responses to complex acoustical stimuli, the present study may provide important information relative to vocal control based on auditory feedback. In previous studies using the auditory perturbation paradigm, a single acoustical dimension was altered. However, in normal audio-vocal situations such as speech and singing, perturbations in both dimensions occur. Many of these are done purposefully such as for suprasegmental contrasts in speech. During normal speech, emphasized syllables generally have a higher F_0, longer duration and greater amplitude (Alain, 1993; Cooper *et al.*, 1985; Eady and Cooper, 1986), although the amplitude changes are generally not as strong as the frequency or durational modifications (Xu and Xu, 2005). Nevertheless, some degree of positive correlation between the changes in these two dimensions is common. However, to our knowledge there is no information on the correlation of unplanned perturbations in these two dimensions during speech. Certainly, vocal jitter and shimmer seem to occur during all vocalizations, but whether or not they are temporally correlated has not been examined. Similarly, slower perturbations as in tremor are likely to co-occur. If such perturbations co-occur during speech and singing, results of the present study suggest that the audio-vocal system is capable of simultaneous stabilization in both acoustical dimensions. This is true if the perturbations are positively or negatively correlated. The only caveat to this interpretation is that there may be greater numbers of errors ("following" responses) in the stabilization process when the two types of perturbations are in conflict with each other. Such errors in vocal stabilization could lead to greater variability in vocal control and the perception of an unsteady voice in some people.

V. CONCLUSION

The results of this study show that when people sustain vowel sounds, neural mechanisms respond to fluctuations in their voice pitch and loudness feedback by producing compensatory responses in voice F_0 and amplitude. These responses are part of a negative feedback loop that functions to stabilize F_0 and amplitude around an actual or intended goal. This study shows that the mechanisms for voice F_0 and amplitude stabilization are predominantly independent. Namely, the system is capable of stabilizing F_0 regardless of whether or not there is a change in loudness feedback, and vice versa. Further, a simple mathematical model using negative feedback, without any dimensional cross coupling or nonlinear interaction, reproduces the main features of our experimental data. In some cases, however, responses are cross-coupled, delayed or exhibit reduced magnitudes if there is more than one type of feedback perturbation, or if the directions of the multiple feedback dimensions are in conflict with each other. These observations suggest that some elements of either the perceptual or compensatory mechanisms interact to a small extent.

Alain, C. (**1993**). "The relation among fundamental frequency, intensity, and duration varies with accentuation," J. Acoust. Soc. Am. **94**, 2434–2436.

Bauer, J. J. (**2004**). "Task dependent modulation of voice F0 responses elicited by perturbations in pitch of auditory feedback during English speech and sustained vowels." Ph.D. dissertation, Northwestern University, Evanston, IL.

Bauer, J. J., and Larson, C. R. (**2003**). "Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," J. Acoust. Soc. Am. **114**, 1048–1054.

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (**2006**). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," J. Acoust. Soc. Am. **119**, 2363–2371.

Binnie, C. A., Daniloff, R. G., and Buckingham, H. W. (**1982**). "Phonetic disintegration in a five-year-old following sudden hearing loss," J. Speech Hear Disord. **47**, 181–189.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 Responses to Manipulations in Pitch Feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063.

Cooper, W. E., Eady, S. J., and Mueller, P. R. (**1985**). "Acoustical aspects of contrastive stress in question-answer contexts," J. Acoust. Soc. Am. **77**, 2142–2156.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of

frequency-shifted auditory feedback on voice F_0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

Eady, S. J., and Cooper, W. E. (**1986**). "Speech intonation and focus location in matched statements and questions," J. Acoust. Soc. Am. **80**, 402–416.

Elliott, L., and Niemoeller, A. (**1970**). "The role of hearing in controlling voice fundamental frequency," Int. Aud. **IX**, 47–52.

Gandevia, S. C., and Rothwell, J. C. (**1987**). "Short-latency cortical projection to human phrenic motoneurons," in *Respiratory Muscles and their Neuromotor Control*, edited by G. C. Sieck, S. C. Gandevia, and W. E. Cameron (Alan R. Liss, New York), pp. 115–119.

Goldstein, J. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.

Hain, T. C., Burnett, T. A., Larson, C. R., and Kiran, S. (**2001**). "Effects of delayed auditory feedback (DAF) on the pitch-shift reflex," J. Acoust. Soc. Am. **109**, 2146–2152.

Heinks-Maldonado, T. H., and Houde, J. F. (**2005**). "Compensatory responses to brief perturbations of speech amplitude," ARLO **6**, 131–137.

Houtsma, A. (**1995**). "Pitch perception," in *Hearing: Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego), pp. 267–291.

Howard, M. A., Volkov, I. O., Abbas, P. J., Damasio, H., Ollendieck, M. C., and Granner, M. A. (**1996**). "A chronic microelectrode investigation of the tonotopic organization of human auditory cortex," Brain Res. **724**, 260–264.

Jaskowski, P., Rybarczyk, K., and Jaroszyk, F. (**1994**). "The relationship between latency of auditory evoked potentials, simple reaction time, and stimulus intensity," Psychol. Res. **56**, 59–65.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. 108, **1246–1251**.

Jones, J. A., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.

Kawahara, H., Hirai, T., and Honda, K. (**1993**). "Laryngeal muscular control under transformed auditory feedback with pitch perturbation," Report No. SP93–39, The Institute of Electronics, Information and Communication Engineers, Japan.

Kawahara, H., and Williams, J. C. (**1996**). "Effects of auditory feedback on voice pitch trajectories: Characteristic responses to pitch perturbations," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, Sydney), pp. 263–278.

Kiran, S., and Larson, C. R. (**2001**). "Effect of duration of pitch-shifted feedback on vocal responses in Parkinson's Disease patients and normal controls," J. Speech Lang. Hear. Res. **44**, 975–987.

Klug, A., Khan, A., Burger, R. M., Bauer, E. E., Hurley, L. M., Yang, L., Grothe, B., Halvorsen, M. B., and Park, T. J. (**2000**). "Latency as a function of intensity in auditory neurons: influences of central processing," Hear. Res. **148**, 107–123.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Larson, C. R. (**1998**). "Cross-modality influences in speech motor control: The use of pitch shifting for the study of F0 control," J. Commun. Dis. **31**, 489–503.

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). "Comparisons of voice F_0 responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am. **110**, 2845–2848.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). "Effects of pitch-shift onset velocity on voice F0 responses," J. Acoust. Soc. Am. **107**, 559–564.

Ludlow, C. L., and Lou, G. (**1996**). "Observations on human laryngeal muscle control," in *Vocal Fold Physiology: Conrolling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 201–218.

McBeath, M. K., and Neuhoff, J. G. (**2002**). "The Doppler effect is not what you think it is: dramatic pitch change due to dynamic intensity change," Psychon. Bull. Rev. **9**, 306–313.

Moore, B. C. J. (**1995**). "Frequency analysis and Masking," in *Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego), pp. 161–205.

Moore, B. C. J., and Sek, A. (**1998**). "Discrimination of frequency glides with superimposed random glides in level," J. Acoust. Soc. Am. **104**, 411–421.

Natke, U., Donath, T. M., and Kalveram, K. T. (**2003**). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.

Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.

Neuhoff, J. G., Kramer, G., and Wayand, J. (**2002**). "Pitch and loudness interact in auditory displays: Can the data get lost in the map?," J. Exp. Psychol., Appl. **8**, 17–25.

Neuhoff, J. G., McBeath, M. K., and Wanzie, W. C. (**1999**). "Dynamic frequency change influences loudness perception: a central, analytic process," J. Exp. Psychol. Hum. Percept. Perform. **25**, 1050–1059.

Palmer, A. R. (**1995**). "Neural signal processing," in *Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego), pp. 75–113.

Sapir, S., McClean, M. D., and Larson, C. R. (**1983**). "Human laryngeal responses to auditory stimulation," J. Acoust. Soc. Am. **73**, 315–321.

Schouten, J. (**1940**). "The residue and the mechanism of hearing," *Proceedings of the Koninklijke Akademie van Wetenschap*.

Seitz, P. F., and Rakerd, B. (**1997**). "Auditory stimulus intensity and reaction time in listeners with longstanding sensorineural hearing loss," Ear Hear. **18**, 502–512.

Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (**2005**). "Voice responses to changes in pitch of voice or tone auditory feedback," J. Acoust. Soc. Am. **117**, 850–857.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A., III. (**1999**). "Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex," J. Neurophysiol. **82**, 2346–2357.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ).

Ward, D., and Burns, E. (**1978**). "Singing without auditory feedback," J. Res. Sing. **1**, 4–44.

Xu, Y., Larson, C., Bauer, J., and Hain, T. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Xu, Y., and Xu, C. X. (**2005**). "Phonetic realization of focus in English declarative intonation," J. Phonetics **33**, 159–197.

The minimum glottal airflow to initiate vocal fold oscillation

Jack J. Jiang[a] and Chao Tao

Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Medical School, Madison, Wisconsin 53792-7375

Phonation threshold flow (PTF) is proposed as a new aerodynamic parameter of the speech production system in this study. PTF is defined as the minimum airflow that can initiate stable vocal fold vibration. Because the glottal airflow can be noninvasively measured, it is suggested that the aerodynamic parameter PTF may be more practical for clinical vocal disease assessment. In order to investigate the relationship between PTF and phonatory system properties, the stability of the body-cover vocal fold model was analyzed. The study has theoretically shown that PTF is a sensitive aerodynamic parameter dependent on tissue properties, glottal configuration, and vocal tract loading. It was predicted that PTF can be reduced by decreasing tissue viscosity, decreasing mucosal wave velocity, increasing vocal fold thickness, or decreasing prephonatory glottal area. Furthermore, it was predicted that a divergent glottis or low vocal tract resistance lead to a reduced PTF. Also discussed is the potential significance of PTF in investigating the energy distribution in a vocal fold vibration system and related clinical applications. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2710961]

PACS number(s): 43.70.Bk, 43.75.Rs [BHS]   Pages: 2873–2881

I. INTRODUCTION

In recent years, researchers have paid much attention to the aerodynamic inputs involved in vocal function because it is believed that these aerodynamic parameters can provide considerable insight into speech system dysfunctions (Holmberg, 2003). Subglottal pressure is one important aerodynamic parameter (Baken and Orlikoff 2000). The phonation threshold pressure (PTP), the minimum subglottal pressure required to initiate phonation, is directly related to tissue properties and glottal configuration. It was predicted that PTP is reduced by reducing the mucosal wave velocity, viscosity, glottal width, and the convergence angle in the glottis (Titze, 1988; 1989; 1992). Lucero (1996) analytically suggested the existence of an optimal glottal width for ease of phonation by including the air pressure losses due to glottal viscous resistance. The differences observed at vocal fold oscillation onset versus offset may be described through the phenomenon of oscillation hysteresis (Lucero, 1999); the oscillation conditions of the vocal folds become more restricted as laryngeal size is reduced (Lucero and Koenig, 2005). The relationship between PTP and vocal fold physiological parameters was also studied by using excised larynx experiments (Jiang *et al.*, 1999a; Verdolini *et al.*, 2002) and physical models (Titze *et al.*, 1995; Chan *et al.*, 1997). These studies showed that PTP is sensitive to the physiological changes of the vocal folds. Vocal fold pathologies, such as dehydration, increase the PTP value (Verdolini *et al.*, 2002). Therefore, it was suggested that the data of PTP could be of great assistance in improving the precision of initial diagnosis for vocal pathologies. However, some problems prevent the broad clinical application of PTP. Most methods to measure PTP are impractical for routine clinical use because they are invasive (Isshiki, 1964; Koike and Perkins, 1968; Kitzing and Lofqvist, 1975). Although the airflow interruption techniques provide the possibility for measuring subglottal pressure noninvasively (Bard *et al.*, 1992; Jiang *et al.*, 1999b; Jiang *et al.*, 2006), these methods usually require thatsubjects maintain constant glottal configuration and constant subglottal pressure even throughout the airflow interruption process. This is usually difficult for untrained subjects (Fisher and Swank, 1997).

Glottal airflow is another important aerodynamic parameter that is associated with laryngeal function and speech production (Baken and Orlikoff, 2000). Glottal airflow is also called the volume velocity. Glottal airflow and subglottal pressure together provide a more comprehensive description of glottal aerodynamics. For a constant subglottal pressure, airflow indicates the impedance characteristics of the airway created by the glottis and the vocal tract. Therefore, airflow data may also be expected to indicate the degree of laryngeal dysfunction just as subglottal pressure might (Rammage *et al.*, 1992; Woodson *et al.*, 1996). Furthermore, empirical results from excised larynx experiments and clinical observations have suggested that glottal airflow should also be above a minimum value in order to initiate vocal fold vibration. This minimum airflow is expected to be related to the tissue properties and glottal configuration. Therefore, similar to PTP data, airflow data under critical phonation conditions might also be helpful in assessing laryngeal pathologies. However, the theoretical explanations of the above mentioned experiences from excised larynx experiments and clinical observation have not been given. Moreover, it is unclear what factors influence the minimum airflow for initiating phonation.

The purpose of this paper is to propose a new aerodynamic parameter, phonation threshold flow (PTF), which is

---
[a]Author to whom correspondance should be addressed. Electronic mail:jjjiang@wisc.edu
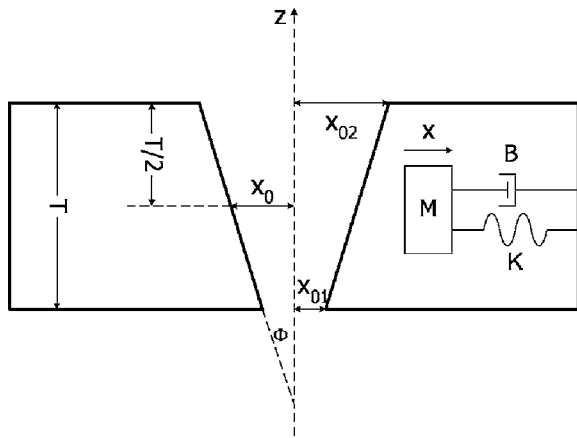
FIG. 1. Sketch map of vocal fold configuration.

defined as the minimum glottal airflow required to initiate phonation. The work in this study is based on the use of a body-cover model of vocal fold structure (Titze, 1988). In this model, the viscoelastic tissue properties are simplified to a mass-spring model with damping. The mucosal wave propagation of the vocal fold cover is represented by a linearized glottal area with a time delay in movement from the bottom to the top of the medial vocal fold surface. The aerodynamics of the glottis is described by a modified Bernoulli equation, and vocal tract loading is characterized by a resistance constant and an inertial constant. This model has been broadly used to study vocal fold vibration because it has a good balance between physiological completeness and mathematic model simplicity (Lucero, 1995). In this paper, we analyze the stability of the body-cover model and propose the critical condition for initiating phonation. We then obtain PTF based on a stability analysis. Last, we discuss potential clinical applications of the new aerodynamic parameter PTF.

II. MODEL AND STABILITY ANALYSIS

The simplified vocal fold model is given in Fig. 1 for the following analysis. The viscoelastic tissue mechanics of the vocal folds can be approached by a simple mechanical oscillator with a spring, a mass block, and a damper

$$M\dot{x} + B\dot{x} + Kx = P_g(x, \dot{x}, t),\tag{1}$$

where M, B, and K are the effective mass, damping and stiffness per unit area of the vocal fold; x is the displacement of the point on the glottal surface at the $T/2$ location. The right term, P_g, is the mean intraglottal pressure applied to the vocal fold surface due to the interaction between the vocal fold and the airflow through glottis.

With the assumption of a linear glottis, the mean intraglottal pressure can be written as (Titze, 1988):

$$P_g = P_i + P_{k_2}(1 - k_e - a_2/a_1),\tag{2}$$

where k_e is a pressure recovery coefficient for the turbulent region downstream of the glottis, and $a_1 = 2L(x_{01} + x + \tau\dot{x})$ and $a_2 = 2L(x_{02} + x - \tau\dot{x})$ are the glottal area at entry and exit, respectively; $2\tau = T/c$ is the time delay from glottal entry to glottal exit due to mucosal wave propagation. L and T are the vocal fold length and thickness, respectively, c is

the mucosal wave velocity. The mucosal wave velocity c is directly related to $(\mu/\rho_m)^{1/2}$, where μ is the transverse shear modulus of the mucosa and ρ_m is the density of the mucosa (Titze, 1976). The terms x_{01} and x_{02} are prephonatory half-glottal widths at entry and exit, respectively. P_i, is the supraglottal pressure (input pressure to the vocal tract). Since the input impedance of the vocal tract is mainly inertial whenever the fundamental frequency is less than the frequency of the first resonance of the vocal tract (Rothenberg, 1981), the input pressure P_i, of the vocal tract can be written as

$$P_i = R_2 U + I_2 \dot{U},\tag{3}$$

where U is the volume velocity of the air flow, and R_2 and I_2 are the resistance and the inertial constant of the vocal tract, respectively. P_{k_2} is the kinetic pressure at glottal exit, which is related to the airflow at the glottal exit by

$$P_{k_2} = (\rho/2)(U/a_2)^2,\tag{4}$$

where ρ is the air density. The subglottal pressure P_s, the kinetic pressure P_{k2}, and the input pressure P_i have the following relationship:

$$P_s - P_i = k_t P_{k2},\tag{5}$$

where k_t, is the transglottal pressure coefficient.

Substituting Eqs. (2)–(4) into Eq. (1) and Eq. (5), a closed dynamic system can be formed to describe the vocal fold vibration:

$$M\dot{x} + B\dot{x} + Kx = R_2 U + I_2 \dot{U} + (\rho/2)(U/a_2)^2(1 - k_e - a_2/a_1).\tag{6a}$$

$$P_s - R_2 U - I_2 \dot{U} = k_t(\rho/2)(U/a_2)^2.\tag{6b}$$

The above nonlinear differential equations are autonomous. This is the one-mass model presented by Titze (1988). In this model, the mucosal wave is imposed instead of the rotational component (Titze, 2002; Titze and Story, 2002) to drive the vocal fold oscillation. Titze (1988) has theoretically predicted the PTP using this model. Experiments (Chan and Titze, 2006) have demonstrated the validity of Titze's theory, which indicates that this model is a useful tool for understanding the critical condition for initiating phonation. Therefore, we also employ this model in this study to introduce the concept of phonation threshold flow. In Eq. (6a), the terms on the right hand side describe the intraglottal pressure applied to the vocal fold medial surfaces due to vocal fold and airflow interaction, which is the driving force of the one-mass oscillatory model. The value of the airflow is solved from Eq. (6b). However, Eq. (6b) also implicitly contains the variables of tissue displacement and velocity, which are solved from Eq. (6a). Therefore, only by solving Eqs. (6a) and (6b) simultaneously can the dynamic characteristics of the vocal fold with vocal tract loading be obtained.

It is difficult to analytically solve Eqs. (6a) and (6b) because they are both nonlinear. In order to get an approximate analytic description for phonatory critical condition, some assumptions have been made to simplify Eqs. (6a) and
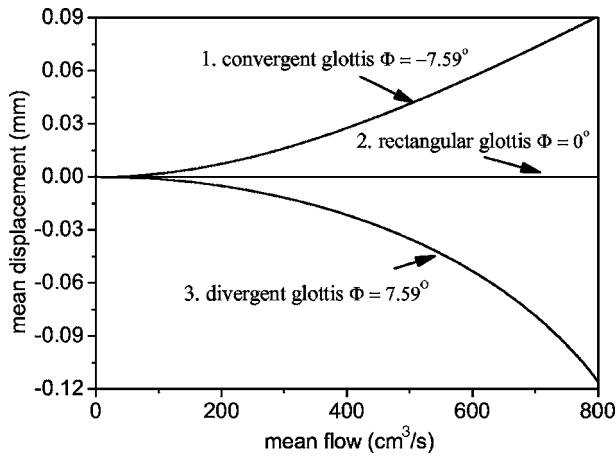
FIG. 2. The mean displacement $\bar{x}$ as a function of the mean airflow $\bar{U}$.

(6b). We decompose U and x into a mean value and a small oscillatory component, that is, $U = \bar{U} + \widetilde{U}$ and $x = \bar{x} + \widetilde{x}$, where $\bar{U}$ and $\bar{x}$ are the mean values of U and x over the cycle, respectively, and $\widetilde{U}$ and $\widetilde{x}$ represent the small oscillatory components. It was also assumed that the oscillatory component satisfies the small-amplitude oscillation condition (Titze, 1988), i.e., $(x_0 + \bar{x})^2 \gg (\widetilde{x} + \tau\dot{\widetilde{x}})^2$. Second, for the sake of simplicity, the vocal tract loading is ignored in the following analysis, that is, $R_2 = 0$ and $I_2 = 0$. Usually, a single mass model in the absence of a vocal fold tract load will not oscillate. However, a mucosal wave velocity is imposed in this model. Therefore, this one-mass model can oscillate even in the absence of a vocal fold tract. According to Eq. (6b), the oscillatory component $\widetilde{U}$ only depends on $\widetilde{a}_2$ and the airflow can be rewritten as

$$U = \bar{U} + (\widetilde{x} - \tau\dot{\widetilde{x}})\bar{U}/(x_{02} + \bar{x}). \qquad (7)$$

With the above assumptions, the simultaneous Eqs. (6a) and (6b) can be decoupled. Substituting Eq. (7) and $x = \bar{x} + \widetilde{x}$ into Eqs. (6a) and (6b) and letting $\ddot{\widetilde{x}} = \dot{\widetilde{x}} = \widetilde{x} = 0$, we have

$$\bar{x} = 0 \qquad (8a)$$

for $x_{01} = x_{02}$; and,

$$K\bar{x}(x_{01} + \bar{x})(x_{02} + \bar{x})^2 = (\rho/8)(\bar{U}/L)^2(x_{01} - x_{02}) \qquad (8b)$$

for $x_{01} \neq x_{02}$. For a divergent glottis ($x_{01} < x_{02}$), the right term of Eq. (8b) is negative. Equation (8b) has two real roots (see Appendix), which are within the open interval $(-x_{01}, 0)$. However, only one of the two real roots corresponds to the vibration (see Appendix). For a convergent glottis ($x_{01} > x_{02}$), the right term of Eq. (8b) is positive. The only physical root is within the interval $(0, +\infty)$ (see Appendix). Figure 2 shows the relationship between mean displacement and mean airflow. Curve 1 corresponds to a convergent glottis with $x_{01} = 0.12$ cm and $x_{02} = 0.08$ cm ($\Phi = 7.59°$). Curve 2 corresponds to a rectangular glottis ($\Phi = 0°$). Curve 3 corresponds to a divergent glottis with $x_{01} = 0.08$ cm and $x_{02} = 0.12$ cm ($\Phi = -7.59°$). The other parameters are $K = 200$ kdyn/cm^3, $\rho = 0.001\,14$ g/cm^3, and $L = 1.4$ cm (Titze, 1988). The con-

vergent glottis corresponds to a positive mean displacement, whereas the divergent glottis corresponds to a negative mean displacement due to Bernoulli's force.

Substituting Eq. (7) into Eq. (6a) and expanding the final term of Eq. (6a) with a Taylor series ignoring the high-order terms, Eq. (6a) can be linearized as

$$\dot{\widetilde{x}} = \widetilde{v};$$

$$\dot{\widetilde{v}} = -(K^*/M^*)\widetilde{x} - (B^*/M^*)\widetilde{v}, \qquad (9)$$

where

$$M^* = M, \qquad (10)$$

$$B^* = B - (\rho/2)(\bar{x}_{01} + \bar{x}_{02})\tau\bar{U}^2/(2L\bar{x}_{01}\bar{x}_{02})^2, \qquad (11)$$

$$K^* = K + (\rho/2)(\bar{x}_{01} - \bar{x}_{02})\bar{U}^2/(2L\bar{x}_{01}\bar{x}_{02})^2. \qquad (12)$$

Also, $\bar{x}_{01} = x_{01} + \bar{x}$ and $\bar{x}_{02} = x_{02} + \bar{x}$. The fixed point of Eq. (9) is $(0,0)$ after the above substitution. It was previously shown that the subglottal pressure can be expressed by the airflow according to the airflow-pressure relationship [Eqs. (2)–(4)]. Therefore, after substituting the airflow-pressure relationship [Eqs. (2)–(4)] into Eqs. (1) and (5), the obtained Eqs. (6a) and (6b) will only explicitly contain the airflow U. Then, based on the small amplitude oscillation assumption, Eq. (9) is obtained where the mean component $\bar{U}$ of airflow is a constant. That is, it is not dependent on the variables $\widetilde{x}$ and $\widetilde{v}$. In this equation, instead of the subglottal pressure, the mean airflow is considered as a control parameter to analyze the dynamic behavior of the vocal folds and obtain the minimum airflow to initiate phonation. The stability of Eq. (9) can be quantitatively analyzed by solving its characteristic equations $|\mathbf{J} - \lambda\mathbf{I}| = 0$, where $\mathbf{J}$ is the Jacobian matrix of Eq. (9), λ represents the characteristic roots, and $\mathbf{I}$ is the unit matrix:

$$\begin{vmatrix} 0 - \lambda & 1 \\ -K^*/M^* & -B^*/M^* - \lambda \end{vmatrix} = \lambda^2 + (B^*/M^*)\lambda + (K^*/M^*)$$

$$= 0. \qquad (13)$$

The above quadratic equation has different roots for different coefficients M^*, B^*, and K^*. Figure 3 displays the regions partitioned in the (B^*/M^*)-(K^*/M^*) parameter space, where the coefficient space is divided into six regions by the coordinate axis and the parabolic curve $(B^*/M^*)^2 = 4(K^*/M^*)$. In each region, Eq. (9) has different local stability qualities at the fixed point $(0, 0)$ (i.e., the vocal fold has different dynamic characteristics). In the regions I and II, where $(K^*/M^*) < 0$, one root of the above quadratic equation is a positive real number and another is a negative real number. The fixed point is a saddle point. Equation (9) cannot generate an oscillation solution for a linear small-amplitude oscillatory system. In region III, where $(B^*/M^*) > 0$ and $(B^*/M^*)^2/4 \geq (K^*/M^*) \geq 0$, one root of Eq. (13) is a negative real number and another root is zero or a negative real number. Therefore, the fixed point in this region is a stable node. The displacement $\widetilde{x}$ solved using Eq. (9) will approximately approach zero for $t \to 0$. In region IV, where $(B^*/M^*) > 0$ and $(B^*/M^*)^2/4 < (K^*/M^*)$, both roots of the quadratic Eq.
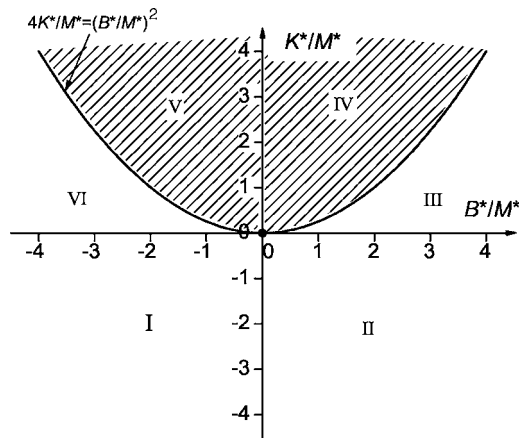
FIG. 3. The distribution of qualitatively different fixed point local stability in (B^*/M^*)-(K^*/M^*) parameters space. I. Saddle; II. Saddle; III. Stable node; IV. Stable focus; V. Unstable focus; VI. Unstable node.

(13) are complex numbers with negative real parts and non-zero imaginary parts. The fixed point is a stable focus and Eq. (9) produces a damping oscillatory solution. In region V, where $B^*/M^* < 0$ and $(B^*/M^*)^2/4 < (K^*/M^*)$, both roots of Eq. (13) are also complex numbers, but their real parts are positive, which means that the fixed point is an unstable focus. The solution of Eq. (9) is a growing oscillatory solution. In region VI, where $B^*/M^* < 0$ and $(B^*/M^*)^2 \geq 4K^*$, both roots of Eq. (13) are positive real numbers. The fixed point is an unstable node. In this case, the output of Eq. (9) is divergent.

In summary, only within the regions IV and V of the (B^*/M^*)-(K^*/M^*) space is an oscillatory output produced. Therefore, these two regions are the domains of interest. In region IV, although the oscillatory output can be generated, the oscillatory amplitude is faded down until its amplitude approaches zero for $t \rightarrow \infty$. The oscillatory output of Eq. (9) will grow in region V, and is the parameter region for phonation. At the edge between region IV and the region V, the vocal fold vibration can just be initiated. Therefore, this edge, $B^*/M^* = 0$ and $K^*/M^* > 0$, is the critical condition for initiating phonation. Because $M^* > 0$ can always be satisfied, the critical condition for initiating phonation is reduced

$$B^* = 0 \quad \text{and} \quad K^* > 0. \tag{14}$$

III. PHONATION THRESHOLD FLOW

The last section showed that the necessary and sufficient condition of initiating phonation is $B^*/M^* \leq 0$ and $(B^*/M^*)^2 < 4(K^*/M^*)$. Solving the inequality $B^*/M^* \leq 0$, the airflow condition of phonation can be obtained

$$\bar{U} \geq U_{PTF}. \tag{15}$$

U_{PTF} is defined as the phonation threshold flow (PTF), which is the minimum value of airflow that initiates vocal fold vibration. U_{PTF} is the positive real root of the equation $B^* = 0$. In this section, we will discuss the influence of glottal configuration, tissue properties, and vocal tract loading on PTF value.

A. Influence of glottal configuration and tissue properties on PTF

Substituting $\tau = T/2c$ into the following equation,

$$B - (\rho/2)(\bar{x}_{01} + \bar{x}_{02})\tau \bar{U}^2/(2L\bar{x}_{01}\bar{x}_{02})^2 = 0 \tag{16}$$

the phonation threshold flow without vocal tract loading is the positive real root of Eq. (16):

$$U_{PTF} = 4L\bar{x}_{01}\bar{x}_{02}\sqrt{\frac{Bc}{(\bar{x}_{01} + \bar{x}_{02})T\rho}}. \tag{17}$$

PTF depends on the glottal shape. Even if the glottis has a constant mean glottal width, different glottal shapes correspond to different PTF values. The glottal angle for a linear glottis (see Fig. 1) was defined to describe the influence of glottal shape on PTF. Positive and negative glottal angles indicate a divergent glottis and a convergent glottis, respectively. A zero glottal angle represents a rectangular glottis. The prephonatory glottal half width at entry and at exit can be rewritten as

$$x_{01} = x_0 - \frac{1}{2}T_2\tan(\Phi), \tag{18}$$

$$x_{02} = x_0 - \frac{1}{2}T_2\tan(\Phi), \tag{19}$$

where $x_0 = (x_{01} + x_{02})/2$ is the mean glottal width.

Figure 4(a) presents PTF as a function of glottal angle, where the solid line represents the numerical solution from Eqs. (6a) and (6b) and the dashed line is predicted by Eq. (17). Figure 4(b) and 4(c) shows the K^* value and mean of x when phonation is initiated, respectively. The input tissue and airflow parameters in Eqs. (6a) and (6b) are $K = 200$ kdyn/cm^3, $M = 0.476$ g/cm^3, $\rho = 0.001\,14$ g/cm^3, $k_t = 1.1$, $x_0 = 0.1$ cm, $L = 1.4$ cm, $T = 0.3$ cm, (Titze, 1988) and $c = 100$ cm/s, (Baer, 1975). In addition, the damping parameter B often associates with some vocal disease, such as a vocal fold scar (Thibeault et al., 2002; 2003). Therefore, three B values (50, 123, and 200 dyn s/cm^3) are chosen for the vocal foldtissue to represent possible pathological conditions. For $B = 123$ dyn s/cm^3 and $B = 200$ dyn s/cm^3, the increase of glottal angle monotonically decreases the PTF value. This is because the mean displacement of a divergent glottis is negative but that of a convergent glottis is positive [see Fig. 4(c)]. The prephonatory width x_0 of all cases in this figure is 0.1 cm. Therefore, the glottal width of a divergent glottis is narrower than that of a convergent glottis even if the prephonatory width is identical. For $B = 50$ dyn s/cm^3, the relationship between PTF and glottal angle is nonmonotonic. PTF increases with an increase in the glottal angle when $\Phi < -7°$. A further increase of the glottal angle will then reduce PTF. In such a condition, the mean displacement $\bar{x}$ is much smaller than the prephonatory glottal width [see Fig. 4(c)], i.e., $|\bar{x}| \ll |x_{01}|, |x_{02}|$. Therefore, substituting Eqs. (18) and (19) into Eq. (17) and ignoring the mean displacement $\bar{x}$, Eq. (17) can be rewritten as:

FIG. 5. PTF as a function of damping per unit area B.
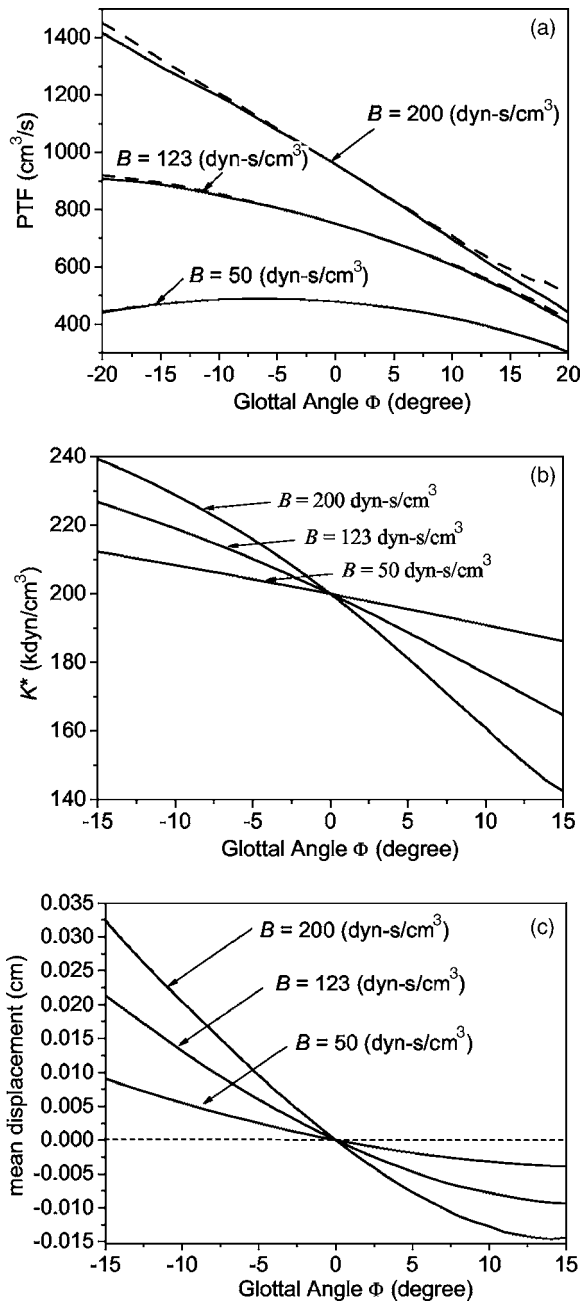
FIG. 4. The influence of glottal configuration on initiating phonation. (a) PTF as the function of glottal angle Φ. (b) K^* as the function of glottal angle Φ when phonation is just initiated, (c) The relationship between mean displacement $\bar{x}$ and glottal angle Φ when phonation is just sustained.

$$U_{\mathrm{PTF}} \approx L(4x_0^2 - T^2\tan^2(\Phi))\sqrt{\frac{Bc}{2x_0 T\rho}}. \qquad (20)$$

Equation (20) implies that PTF should have a maximum value near $\Phi=0$ in this condition.

PTF is associated with the glottal size. Figure 4(c) has shown that the mean displacement $\bar{x}\to 0$ ($\bar{x}_{01}\approx\bar{x}_{02}\approx x_0$) when $\Phi\to 0$. Therefore, for a rectangular glottis or a glottis with small glottal angle, Eq. (17) can be simplified to

$$U_{\mathrm{PTF}} = G_A\sqrt{\frac{Bc}{e\rho}}, \qquad (21)$$

where $G_A = 2x_0L$ is the glottal area and $e = T/2x_0$ is the axial duct aspect ratio of the rectangular glottis. Equation (21)
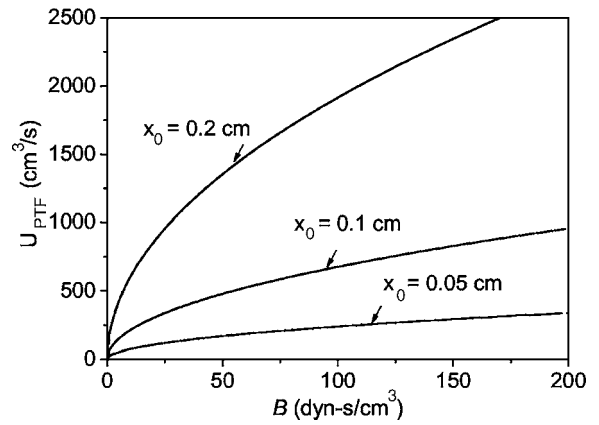
suggests that an axially longer and narrow glottis or a small glottal area will decrease PTF.

PTF also depends on the tissue properties of the vocal fold. Equation (21) predicts that PTF is reduced by decreasing tissue viscosity B and mucosal wave velocity c. Titze (1976) has theoretically shown that the mucosal wave velocity is directly proportional to the square root of the transverse shear modulus and inversely proportional to the square root of the tissue density. Coupling this theoretical result with Eq. (21), PTF is reduced by decreasing tissue viscosity and tissue transverse shear modulus. Figure 5 presents PTF as a function of damping per unit area B with the glottal half widths $x_0 = 0.05$, 0.1, and 0.2 cm, respectively, where the glottis shape is rectangular ($\Phi=0°$) and the other parameters are identical with those of Fig. 4. The solid lines are obtained by numerically solving Eqs. (6a) and (6b). The results predicted by Eq. (21) are also presented with dashed lines. Equation (21) is obtained with a small-amplitude assumption which fits the exact solution of Eqs. (6a) and (6b) very well. Therefore, the dashed lines are almost coincident with the solid lines. It is clear that PTF is increased by increasing tissue damping.

B. Influence of vocal tract loading on PTF

Vocal tract loading is not considered in the above analysis and discussions. However, vocal tract loading can also shift the PTF value. In this subsection, we will study the potential influence of vocal tract loading on PTF by using numerical simulation. For the sake of simplicity, the glottal configuration was assumed to be rectangular ($x_{01}=x_{02}$). Figure 6 presents PTF as the function of vocal tract resistance R_2 and Fig. 7 presents PTF as the function of vocal tract inertia I_2, where the PTF value is obtained by numerically solving the Eqs. (6a) and (6b). In this simulation, the other parameters of vocal fold tissue and glottal size are $K = 200$ kdyn/cm^3, $M=0.476$ g/cm^3, $B=123$ dyn s/cm^3 $\rho = 0.0014$ g/cm^3, $k_t = 1.1$, $x_0 = x_{01} = x_{02} = 0.1$ cm, $L=1.4$ cm, $T=0.3$ cm, and $c=100$ cm/s. It is seen that in our simulation range, PTF is reduced due to an increase in the inertial constant I_2 or a decrease in resistance R_2. Therefore, an increase in vocal tract inertia or a decrease in vocal tract resistance could make phonation easy.
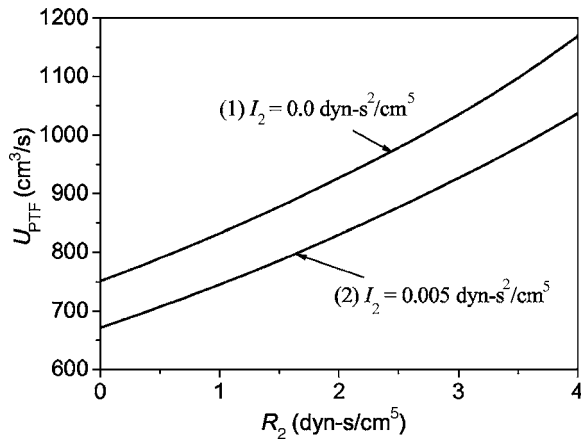
FIG. 6. PTF as a function of vocal tract resistance R_2, where the solid lines are the exact numerical solutions of Eqs. (6a) and (6b).

C. Influence of tissue stiffness (K) on initiating phonation

According to the analysis in Sec. II, the condition $K^* > 0$ must also be satisfied in order to generate an oscillation solution. Therefore, $\bar{U} \geq U_{PTF}$ is just one necessary condition for initiating phonation. Even if airflow is beyond PTF predicted by $B^* = 0$, phonation still could not be initiated if the inequality $K^* > 0$ cannot be satisfied. Figure 4(b) displays K^* as a function of Φ. It is seen that the value for K^* in the above simulation is much higher than zero, and the condition $K^* > 0$ can be satisfied. However, if the tissue stiffness K is abnormally small, dysphonia could occur since the condition $K^* > 0$ will not be satisfied. Figure 8 presents the dynamic characteristics of vocal folds with the glottal angle $\Phi = 15°$. The other parameters are $B = 123$ dyn s/cm^3, $M = 0.476$ g/cm^3, $\rho = 0.0014$ g/cm^3, $k_t = 1.1$, $x_0 = 0.1$ cm, $L = 1.4$ cm, $T = 0.3$ cm, and $c = 100$ cm/s. In Fig. 8(a), the tissue stiffness is $K = 200$ kdyn/cm^3. Trajectory 1 corresponds to the condition in which the airflow is lower than PTF. It is seen that Eqs. (6a) and (6b) have a damping oscillating solution. The vocal folds do not vibrate. Trajectory 2 corresponds to the condition that the airflow just reaches PTF. At this condition, $B^* = 0$ and $K^* > 0$, the solution is oscillatory. Trajectory 3 corresponds to the condition where airflow is
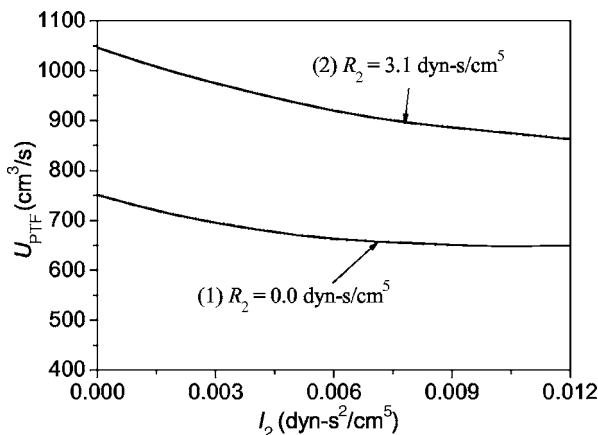
FIG. 7. PTF as a function of vocal tract resistance I_2, where the solid lines are the exact numerical solutions of Eqs. (6a) and (6b).

beyond PTF, where $B^* < 0$ and $4K^*/M^* > (B^*/M^*)^2$. In this case, oscillation grows. However, for an abnormally small value of tissue stiffness, for example $K = 20$ kdyn/cm in Fig. 8(b), the low airflow still generates a stable solution (trajectory 1). With an increase of airflow, K^* will reach zero before B^* reaches zero. Oscillation cannot be produced under this condition [trajectory 2 in Fig. 8(b)]. Further increase of airflow will make $K^* < 0$. The solution here is divergent and oscillation cannot be generated at all [trajectory 3 in Fig. 8(b)]. This theoretical result matches the experimental observation, where the oscillation cannot be generated in an excised larynx that does not build up enough tension.

D. Relationship to other theory

Titze and others have shown that the subglottal pressure must exceed a minimum pressure value in order to initiate vocal fold vibration (Titze, 1988). He defined this minimum pressure as phonation threshold pressure (PTP). This is the critical condition of airflow pressure for initiating phonation. Further experimental studies have supported Titze's theory (Chan and Titze, 2006). Therefore, the one-mass model is a useful tool for studying the critical condition for initiating phonation. This study presents the critical condition of airflow for initiating phonation based on Titze's one-mass model. The subglottal pressure can be associated with the airflow through Eq. (6b). Substituting $U = \bar{U} + \tilde{U}$ and $x = \bar{x} + \tilde{x}$ into Eq. (6b) and ignoring the small-amplitude oscillating terms, we have the relationship between PTF defined in this study and PTP defined by Titze (1988):

$$P_{PTF} \approx R_2 U_{PTF} + k_t(\rho/2)(U_{PTF}/\bar{a}_2)^2, \quad (22)$$

where $\bar{a}_2 = 2L\bar{x}_{02}$. This determinate relationship between PTF and PTP implies that PTF can also provide useful clinical information just as PTP can.

Titze's one-mass model is one of the simplest models of the vocal folds. Many important factors, which could influence the critical condition of initiating phonation, are ignored in this model. Recent works have reported on the influence of these factors on the critical condition by modifying the one-mass model (Lucero, 1996). Lucero (1996) introduced an explicit term for air pressure losses due to glottal viscous resistance into the one-mass model. Using this modified model, Lucero (1996) predicted that there exists an optimal glottal width for ease of phonation. This result agrees with the experimental measurements of Titze *et al.* (1995). The glottal viscous resistance was not included in the one-mass model used in this study. Therefore, the influence of glottal viscous resistance on PTF is not predicted in Eq. (17). However, applying the same small-amplitude oscillatory assumption and the stable analysis on the modified one-mass model, the relationship between the glottal viscous resistance and the PTF can also be obtained.

By analyzing the body-cover model of the vocal folds in the general case of large amplitude vibrations, Lucero (1995) predicted that the minimum lung pressure required to sustain vocal fold oscillation after its onset is lower than the threshold pressure needed to initiate it. In this study, the PTF is defined as the minimum airflow to initiate the phonation.

J. J. Jiang and C. Tao: The minimum glottal airflow for phonation
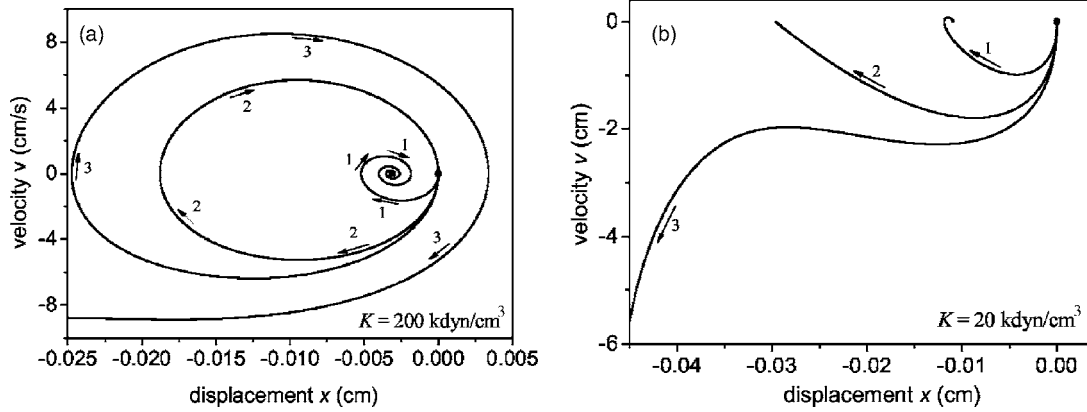
FIG. 8. Trajectories of vocal fold vibration in x-v phase space, where the glottal angle is $\Phi = 15°$. (a) The trajectories of a vocal fold with normal stiffness $K = 200$ kdyn/cm^3, where trajectory 1 (stable focus): $B^* > 0$ and $4K^*/M^* > (B^*/M^*)^2$; trajectory 2: $B^* = 0$ and $K^* > (B^*/M^*)^2$; and trajectory 3 (unstable focus): $B^* < 0$ and $4K^*/M^* > (B^*/M^*)^2$. (b) The trajectories of a vocal fold with abnormal low stiffness $K = 20$ kdyn/cm^3, where trajectory 1: $B^* \approx 108$ dyn s/cm^3 and $K^* \approx 15.4$ kdyn/cm^3; trajectory 2: $B^* = 71$ dyn s/cm^3 and $K^* \approx 0$ kdyn/cm^3; trajectory 2: $K^* < 0$ kdyn/cm^3

Therefore, according to Lucero's theory (1995), it can be expected that the minimum airflow needed to sustain vocal fold oscillation after its onset might also be lower than the PTF value presented in this study. The minimum airflow to sustain vocal fold vibration is needed to be studied in further work.

IV. DISCUSSION AND CONCLUSION

In this study, the general critical conditions for initiating phonation are presented. Based on the stability analysis of the vocal fold model, it was found that the body-cover model generates a damping oscillating solution when $(B^*/M^*)^2 < 4K^*/M^*$ and $B^*/M^* > 0$. However, the same model generates a growing oscillatory solution when $(B^*/M^*)^2 < 4K^*/M^*$ and $B^*/M^* < 0$. Therefore, the critical condition for initiating phonation is the boundary formed by the two regions: $B^* = 0$ and $K^* > 0$. The airflow at the critical condition is defined as the phonation threshold flow, which is the minimum airflow needed to initiate stable vocal fold vibration. This study theoretically shows that PTF can be reduced by reducing the glottal area, tissue viscosity, or mucosal wave velocity, and by increasing the vertical length of the glottal duct, by reducing the prephonatory convergent angle, or by increasing the prephonatory divergent angle. Moreover, vocal tract loading shifts the PTF value. PTF can be reduced by decreasing vocal tract resistance.

The aerodynamic parameter PTF can be useful in understanding vocal fold vibration and the energy transformation in a speech producing system. The vocal fold system plays the role of energy transducer during phonation (Titze, 1994), which distributes the potential energy of the gas in the lungs to the speech producing system. Airflow and air pressure together provide the description of the energy: airflow power can be expressed as the product of the airflow and the pressure. The airflow energy information before phonation is initiated ($\dot{U} = 0$) can be obtained from Eq. (6b):

$$W_{source} = P_s \cdot U = W_{tract} + W_{glottis} + W_{air}, \quad (23)$$

where $W_{tract} = R_2 U^2$ is the power expended due to vocal tract resistance, $W_{air} = (\rho/2)(U^3/\bar{a}_2^2)$ is the airflow kinetic

power at the glottal exit, and $W_{glottis} = (k_t - 1)W_{air}$ is the power expended by the glottal resistance. Because the tracheal area is much larger than the glottal area, the airflow particle velocity in the trachea is much smaller than that at glottal exit and the kinetic power of airflow at the trachea can be ignored in comparison with that at the glottal exit. The term W_{air} can be approximately considered as the increase of airflow kinetic energy after it goes through the glottis. Figure 9 presents the relationship between the airflow and the power distribution in a vocal fold system before phonation has started. It is seen that the source energy distributes into three main parts. One part of the source energy will be transferred to kinetic energy of airflow. The remaining source energy is consumed by glottal resistance $[(k_t - 1)W_{air}]$ and vocal tract resistance. Consequently, the critical condition for initiating phonation can be re-described using the flow energy: PTW = PTP × PTF, where PTW is the phonation threshold power. From the aspect of power transformation, the critical condition for initiating phonation essentially reflects the energy balance change in the vocal fold vibration system. Before the critical power is reached, all the energy will be consumed by
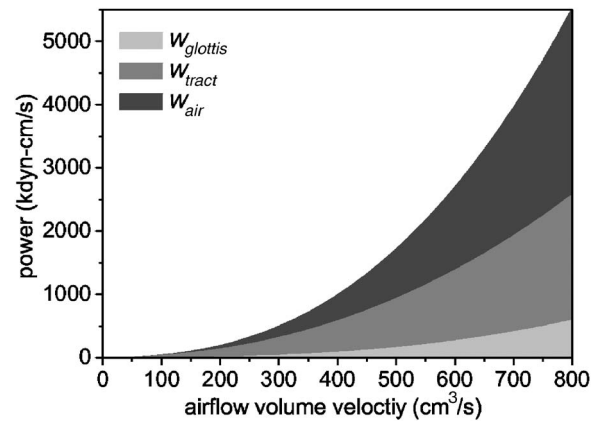
FIG. 9. The energy distribution in the vocal fold system before phonation is started, where $W_{glottis}$ is power consumed by the glottal resistance, W_{air} is the power transferred to the airflow kinetic power, W_{truct} is the power expanded by vocal tract resistance.
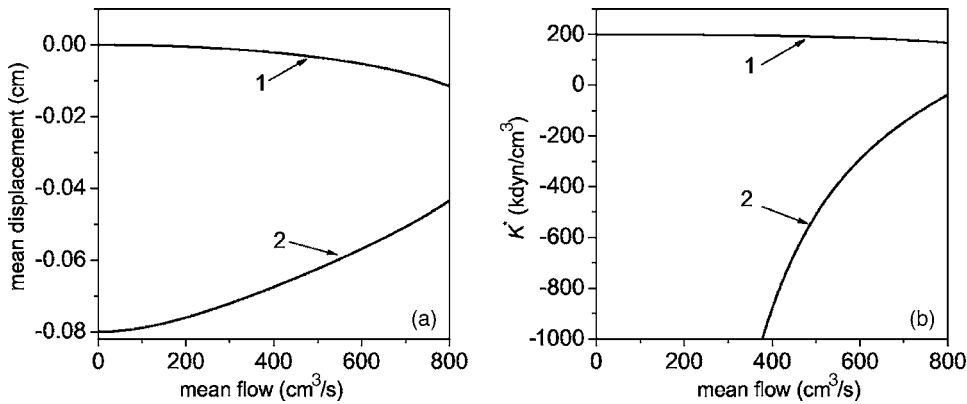
FIG. 10. (a) The two real roots of Eq. (8b) with $x_{01}=0.08$ cm and $x_{02}=0.12$ cm as a function of airflow, (b) The K^* value corresponds to roots 1 and 2 in (a).

acceleration of airflow, glottal resistance, and vocal tract resistance. When the source energy exceeds the critical power, the additional energy will be used to initiate vocal fold oscillation. The dysphonia due to some vocal fold pathologies, such as breathiness and hoarseness, can be partly understood from the concept of flow power. In some cases, the glottal gap is wide. Too much airflow escapes from the glottis. Most of the source energy of these systems is used to accelerate the airflow, while only a small amount is used to initiate the phonation. This inefficient utilization of the source energy accompanies dysphonia in these pathological cases.

The aerodynamic, parameter PTF defined in this paper may also provide a new sensitive indicator for the functional efficiency of the laryngeal system. Clinical studies show that vocal fold disease can change the vocal fold biomechanics and increase the prephonatory glottal width. Paralysis can shift the viscosity and stiffness of vocal tissue, edema alters the glottal shape and size, and vocal scarring due to vocal fold surgery can also change the tissue properties, which are the key elements in determining the minimum airflow for initiating phonation. PTF data under these pathological conditions could be a powerful tool in reflecting the degree of laryngeal dysfunction. Recently, we have done some work to examine the PTF theory in excised larynx experiments. We studied the relationship between vocal fold elongations on PTF and found that PTF is increased with an increase in elongation. We have also studied the influence vocal abduction (prephonatory glottal area) on PTF and PTP. It was found that the PTF is more sensitive to the change of prephonatory glottal area than PTP. The detailed results will be reported in future papers. These works primarily show that PTF can indicate the change of vocal fold properties as well as PTP. Furthermore, the mean airflow from the mouth is roughly equal to the amount exiting the glottis (Baken and Orlikoff, 2000). The glottal airflow can be obtained by measuring the airflow exiting the mouth during phonation. The flow transducer attached to the outlet of an anesthesia-type face mask with appropriate support electronics is commonly used to measure airflow (Baken and Orlikoff, 2000). Using such techniques, valid and reliable aerodynamic measurements can be routinely obtained from children as young as 2 years of age, with a bit of care, skill, and experience (Lotz, 1993). These studies suggest that PTF could be more easily noninvasively evaluated in clinics than other aerodynamic

parameters such as PTP. Therefore, PTF data as an indicator of pathology may be more practical in a clinical setting. This theoretical study of PTF may thus have important clinical potential.

ACKNOWLEDGMENTS

APPENDIX

When $x_{01} \neq x_{02}$, the equilibrium point of the one-mass model can be determined from the bi-quadratic equation [Eq. (8b)]. In the following discussion, $\Psi(\bar{x})$ represents the left term of Eq. (8b). For a convergent glottis ($x_{01} > x_{02}$), the right term of Eq. (8b) is positive. When $x \in [-x01,0], \Psi(\bar{x}) \leq 0$ can always be satisfied; therefore, there are no real roots for Eq. (8b). When $x \in (0, +\infty)$, it is known that $\Psi(0) = 0, \Psi(+\infty) = +\infty$, and $\Psi(\bar{x})$ is a monotonically increasing function. Therefore, there is one real root within this interval according to the intermediate value theorem. When $x \in (-\infty, -x_{01}), \Psi(-x_{01}) = 0, \Psi(-\infty) = +\infty$ and $\Psi(\bar{x})$ is a monotonically decreasing function, there is one real root within $(-\infty, -x_{01})$. However, $\bar{x} < x_{01}$ indicates that the glottis is closed. Therefore, the negative root within $(-\infty, -x_{01})$ is not a physical root and the positive root within $(0, +\infty)$ is the only physical root of Eq. (8b) for a convergent glottis.

For a divergent glottis ($x_{01} < x_{02}$), the right term of Eq. (8b) is negative. However, within $(0, +\infty)$ and $(-\infty, -x_{01}), \Psi(\bar{x}) \geq 0$ can always be satisfied; therefore, there is no real root within this domain. Then, solving $d\Psi(\bar{x})/d\bar{x}=0$, we have three roots:

$$r_1 = -x_{02},$$

$$r_2 = \left[-(3x_{01} + 2x_{02}) + \sqrt{(3x_{01} + 2x_{02})^2 - 16x_{01}x_{02}} \right] / 8,$$

$$r_3 = \left[-(3x_{01} + 2x_{02}) - \sqrt{(3x_{01} + 2x_{02})^2 - 16x_{01}x_{02}} \right] / 8.$$

These three roots correspond to the three extreme points of the function $\Psi(\bar{x})$. Moreover, it is easy to prove that only root r_2 is within the domain $(-x_{01}, 0)$ and $\Psi(r_2)$ is the minimum value of $\Psi(\bar{x})$. Therefore, within $(-x_{01}, 0)$, according to

J. J. Jiang and C. Tao: The minimum glottal airflow for phonation

the intermediate value theorem and $\Psi(0)=\Psi(-x_{01})=0$, there are two real roots when $\Psi(r_2) \leqslant (\rho/8)(\bar{U}/L)^2(x_{01}-x_{02})$.

In addition, although there are two possible $\bar{x}$ values within the open interval $(-x_{01},0)$. Further simulation shows that only one value could correspond to vocal fold vibration of the initiating phonation. Figure 10(a) presents the values of mean displacement with parameter values $K=200$ kdyn/cm^3, $M=0.476$ g/cm^3, $\rho=0.001\,14$ g/cm^3, $L=1.4$ cm (Titze, 1988) and $x_{01}=0.08$ cm, $x_{02}=0.12$ cm. Figure 10(b) gives the corresponding K^* value. It is seen that one root is close to 0 [the curve 1 in Fig. 10(a)] and another is close to $-x_{01}$ [the curve 2 in Fig. 10(a)]. Moreover, the first root corresponds to a positive K^* [the curve 1 in Fig. 10(b)], but the second root always corresponds to a negative K^*. This means that the second fixed point is a saddle point, which does not represent an oscillatory solution in a linear small-amplitude oscillatory system. Therefore, for a divergent glottis, only one fixed point is a focus point, which corresponds to an oscillatory solution of a linear small-amplitude situation.

Baer, T. (1975). "Investigation of phonation using excised larynxes," Ph.D. dissertation, MIT, Cambridge, MA.

Baken, R. J., and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular, Thomson Learning, San Diego).

Bard, M. C., Slavit, D. H., McCaffrey, T. V., and Lipton, R. J. (1992). "Noninvasive technique for estimating subglottic pressure and laryngeal efficiency," Ann. Otol. Rhinol. Laryngol. 101, 578–582.

Chan, R. W., Titze, I. R., and Titze, M. R. (1997). "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. 101(6), 3722–3727.

Chan, R. W., and Titze, I. R. (2006). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. 119(4), 2351–2362.

Fisher, K. V., and Swank, P. R. (1997). "Estimating phonation threshold pressure," J. Speech Lang. Hear. Res. 40, 1122–1129.

Holmberg, E. B., Doyle, P., Perkell, J. S., Hammarberg, B., and Hillman, R. E. (2003). "Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy," J. Voice 17, 269–282.

Isshiki, N. (1964). "Regulatory mechanism of voice intensity variation," J. Speech Hear. Res. 7, 17–29.

Jiang, J. J., Leder, C., and Bichler, A. (2006). "Estimating subglottal pressure using incomplete airflow interruption," Laryngoscope 116, 89–92.

Jiang, J. J., Ng, J., and Hanson, D. (1999a). "The effects of rehydration on phonation in excised canine larynges," J. Voice 13(1), 51–59.

Jiang, J. J., O'Mara, T., Conley, D., and Hanson, D. (1999b). "Phonation threshold pressure measurement during, phonation by airflow interruption," Laryngoscope 109, 425–432.

Kitzing, P., and Lofqvist, A. (1975). "Subglottal and oral air pressure during phonation-preliminary investigation using a miniature transducer system," Med. Biol. Eng. 13, 644–648.

Koike, Y., and Perkins, W. (1968). "Application of miniaturized pressure transducer for experimental speech research," Folia Phoniatr (Basel) 20, 360–380.

Lotz, W. K., D'Antonio, L. L., Chait, D. H, and Netsell, R. W. (1993). "Successful nasoendoscopic and aerodynamic examinations of children with speech/voice disorders," Int. J. Pediatr. Otorhinolaryngol. 26, 165–172.

Lucero, J. C. (1995). "The minimum lung pressure to sustain vocal fold oscillation," J. Acoust. Soc. Am. 98(2), 779–784.

Lucero, J. C. (1996). "Relation between the phonation threshold pressure and the prephonatory glottal width in a rectangular glottis," J. Acoust. Soc. Am. 100(4), 2551–2554.

Lucero, J. C. (1999). "A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset," J. Acoust. Soc. Am. 105(1), 423–431.

Lucero, J. C., and Koenig, L. L. (2005). "Phonation thresholds as a function of laryngeal size in a two-mass model of the vocal folds (L)," J. Acoust. Soc. Am. 118(5), 2798–2801.

Rammage, L. A., Peppard, R. C., and Bless, D. M. (1992). "Aerodynainic, laryngoscopic, and perceptual-acoustic characteristics in dysphonic females with posterior glottal chinks: A retrospective study." J. Voice 6, 64–78.

Rothenberg, M. (1981). "Acoustic Interaction between the Glottal Source and the Vocal Tract," in *Vocal Fold Physiology*, edited by K. Stevens and M. Hirano (University of Tokyo, Japan), pp. 304–323

Thibeault, S. L., Gray, S. D., Bless, D. M., Chan, R. W., and Ford, C. N. (2002). "Histologic and rheologic characterization of vocal fold scarring," J. Voice 16(1), 94–104.

Thibeault, S. L., Bless, D. M., and Gray, S. D. (2003). "Interstitial protein alterations in rabbit vocal fold with scar," J. Voice 17(3), 377–383.

Titze, I. R. (1976). "On the mechanics of vocal-fold vibration," J. Acoust. Soc. Am. 60(6), 1366–1380.

Titze, I. R. (1988). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. 83(4), 1536–1552.

Titze, I. R. (1989). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. 85(2) 901–906.

Titze, I. R. (1992). "Phonation threshold pressure: A missing link in glottal aerodynamics," J. Acoust. Soc. Am. 91, 2926–2935.

Titze, I. R. (1994). *Principles of Voice Production* (Prentice-Hall, Upper Saddle Rivers, NJ).

Titze, I. R., Schmidt, S. S., and Titze, M. R. (1995). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. 97, 3080–3084.

Titze, I. R. (2002). "Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model," J. Acoust. Soc. Am. 111(1), 367–376.

Titze, I. R., and Story, B. H. (2002). "Rules for controlling low-dimensional vocal fold models with muscle activation," J. Acoust. Soc. Am. 112(3), 1064–1076.

Verdolini, K., Min, Y., Titze, I. R., Lemke, J., Brown, K., Mersbergen, M. V., Jiang, J. J., and Fisher, K. (2002). "Biological mechanisms underlying voice changes due to dehydration," J. Speech Lang. Hear. Res. 45, 268–281.

Woodson, G. E., Rosen, C. A., Murry, T., Madasu, R., Wong, F., Hengesteg, A., and Robbins, K. T. (1996). "Assessing vocal function after chemoradiation for advanced laryngeal carcinoma," Arch. Otolaryngol. Head Neck Surg. 122, 858–864.

Perceptual rate normalization in naturally produced rate-varied speech[a)]

Kyoko Nagao[b)] and Kenneth de Jong[c)]
Department of Linguistics, Indiana University, Bloomington, Indiana 47405

The perception of voicing categories is affected by speaking rate, so that listeners' category boundaries on a VOT continuum shift to a lower value when syllable duration decreases [Miller and Volaitis, Percept. Psychophys. **46**, 505–512 (1989); Volaitis and Miller, J. Acoust. Soc. Am. **92**, 723–735 (1992)]. Previous rate normalization effects have been found using artificially varied stimuli. This study examines the effect of speech rate on voicing categorization in naturally produced rate-varied speech. The stimuli contained natural decreases in VOT with faster speech rates so that VOT values for /b/ and /p/ overlapped at the fastest rates. Consonant identification results showed that the rate effects on the perceptual boundary between /p/ and /b/ very closely matched the effects of rate on the productions, though there was a small mismatch with fast rate productions whereby voiced stops were systematically miscategorized as voiceless. Another group of listeners judged the goodness of the consonant, indicating that best exemplars were rate-varied and shifted away from the /p/-/b/ boundary. These results are discussed in light of exemplar-based and abstractionist models of speech perception. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713680]

I. INTRODUCTION

How listeners cope with variability in the speech signal with respect to linguistic categories is a longstanding issue of speech perception. Speaking rate is one major source of such acoustic variability. Speaking rate affects acoustic properties of many kinds, such as consonant duration and vowel duration (Crystal and House, 1988), voice onset time (VOT) for voicing contrasts of syllable initial stops (Miller *et al.*, 1986; Volaitis and Miller, 1992), closure duration for the voicing distinction of intervocalic stops (Lisker, 1957), transition durations for the manner distinction between stops and glides (Miller and Baer, 1983), and durations of adjacent segments [e.g., Repp and Lin (1991) for the preceding fricative /s/; Newman and Sawusch (1996) for the preceding and following consonants]. It is also known that speaking rates vary both across speakers and within speakers (Miller *et al.*, 1984c). Perceptual rate effects are not restricted to local syllable-internal temporal variation. Syllable external tempo estimates also have an effect on the perception of rate-sensitive information (Kidd, 1989; Summerfield, 1981; Wayland *et al.*, 1994; Newman and Sawusch, 1996; 2000; Miller *et al.*, 1984b).

Inter- and intraspeaker variability in speaking rate poses a problem in a theory of speech perception because many phonetic contrasts are differentiated by temporal differences. Unlike traditional views in which individual speech segments correspond to an invariant phonetic representation, the direct realism view (Fowler, 1977, 1980) and the exemplar-based or episodic models of categorization (e.g., Goldinger, 1996; Pisoni, 1997; Nosofsky, 1992) propose that listeners preserve fine phonetic details of their experiences with speech, and this would include contextual speaking rate variation. All of the fine detail is encoded in the perceptual structure that listeners use to categorize the speech. Since all of this information on temporal dynamics is encoded and stored in the long-term memory, perceptual rate normalization should directly match rate variation effects in production.

However, there are cases of mismatch found between speech production and perception as we will show in detail below. The current paper proposes that certain of these cases of mismatch between speech production and perception are partially due to the use of artificially created stimuli in perception studies. More specifically, the current study focuses on the effect of speaking rate on production and perception of English voicing categories.

Among various measures, VOT is a critical acoustic property to differentiate the voicing contrast of syllable initial stops in English (Abramson and Lisker, 1985; Lisker and Abramson, 1964; Lisker, 1975). It has been known that speaking rate affects VOT; VOT increases as syllable duration increases. Furthermore, the rate effect is not symmetrical with respect to the voicing categories. VOT in voiceless consonants changes more with rate than in voiced consonants (Miller and Volaitis, 1989; Volaitis and Miller, 1992). Since VOT is rate sensitive, any model of the perception of linguistic contrasts must account for how listeners are affected by such variability. How can listeners distinguish voicing categories such as /p/ and /b/ in the presence of rate-induced variation in the speech signal?

The most extensive and comprehensive research on VOT and rate sensitivity has been conducted by Joanne Miller and her colleagues over the last two decades. They estimated the VOT values at the voicing category boundary based on production results and perception results. In addition, Miller and her colleagues conducted both production and perception studies to address the nature of voicing categories in English. There is clear evidence that speaking rate affects the location of voicing boundaries. However, close examination of their results also revealed that there is a pervasive mismatch between the perceptual boundary and the production boundary reported in their studies.

Miller *et al.* (1986) explored rate effects in productions of the voicing contrast for initial bilabial consonants in English. They collected the syllables /bi/ and /pi/ produced by three native speakers of English, by employing a magnitude production technique to elicit the syllables across a wide range of speech rates. In this method, the subjects repeat each syllable at a normal rate, and then they repeat the same syllable at a rate proportionally slower or faster than the normal rate. Using the same method with other stops, Volaitis and Miller (1992) replicated these findings and noted that, although the large rate effect on VOT seen in /pi/ becomes smaller when the values were transformed into log scales, the difference between /bi/ and /pi/ remained. Similarly, Kessinger and Blumstein (1997) examined the relationship between VOT and speaking rate and found that the voiceless stops are produced with shorter VOT at fast speaking rates than at slow speaking rates, whereas the VOT for voiced stops does not change at different rates. These asymmetrical rate effects on voicing contrasts have also been found in other points of articulation with /t/-/d/ and /k/-/g/ contrasts (Kessinger and Blumstein, 1997; Volaitis and Miller, 1992) and in other languages (Kessinger and Blumstein, 1997; Magloire and Green, 1999).

Specific results for the production of English initial labial consonants are summarized in Table I. These studies are listed in terms of the rate of speech they examined (fast, normal or medium, and slow), though some of the studies did not report the actual speaking rate. Speakers were usually asked to read the words at a small number of rates, except Miller *et al.* (1986) and Volaitis and Miller (1992), and rates were controlled subjectively using labels such as normal, fast, and slow rates. Even in the most well designed studies by Miller and her colleagues, initial speaking rate (normal) was determined by each individual speaker and other speaking rates were manipulated relative to the initial rate that each speaker employed. Some studies determined speaking rate in terms of subjective rates without specific acoustic measures, and some use either syllable duration or vowel duration as a reference of speaking rates. Even among those studies, there are slight contextual differences in the data collection such as words produced in isolation versus words in a sentence. Also, there are various methodological differences between the studies such as elicitation methods, the following vowel quality, and the linguistic context that each utterance was produced. Still, average VOT values for /b/ are relatively consistent across the studies and appear at around 10 ms across different speaking rates. On the other hand,

average VOT values reported for /p/ are quite different among those studies. Within normal or medium rates, average VOT values span from 28 to 80 ms. In addition, the range of VOT in the productions of /p/ is wider than VOT ranges in /b/ productions across studies.

VOT values at the boundary between /b/ and /p/ were estimated in some of the studies (Miller *et al.* 1986; Miller and Volaitis, 1989; Volaitis and Miller, 1992; Kessinger and Blumstein, 1997). As can be seen in Table I, the estimated VOT values at category boundary between English /b/ and /p/ are also similar across the studies.

In Miller *et al.* (1986), category boundaries for production, again, in English, were estimated in terms of VOT. These optimal VOT boundaries, computed to differentially identify speakers' intended consonant categories, classified the intended category (/b/ or /p/) with more than 90% accuracy. When computed from their entire data set, the optimal VOT boundary was at 23.5 ms. They also computed rate-varied optimal VOT values for data within successive 50-ms syllable duration intervals from 100 to 700 ms. Rate-dependent optimal values improved categorization performance to 97.6%. In other words, allowing the VOT boundary to shorten at fast rates increased classification accuracy.

Such rate effects on the articulation of segments are reflected in perceptual responses to rate-varied stimuli. Miller and Volaitis (1989) examined rate effects on voicing perception by systematically changing VOT values for /bi/ and /pi/. The VOT continuum was created to extend from English /b/ through /p/ to an exaggerated category beyond /p/ (called */p/). To simulate slow and fast speech, two different durations of syllables were synthesized with a serial resonant synthesizer. Listeners were asked to identify each of these stimuli as either /b/, /p/, or an exaggerated */p/. Listeners exhibited later VOT classification boundaries between /b/ and /p/ in long syllables than in short syllables. Volaitis and Miller (1992) extended the study to stops produced at other points of articulation (/g/, /k/, and */k/) and found similar results. Also, similar results of rate effects on the VOT boundary between voiced and voiceless stops have been found by other investigators (Kessinger and Blumstein, 1997; Summerfield, 1981; Sawusch and Newman, 2000).

To probe the internal structure of the listeners' categories, Miller and Volaitis (1989) and Volaitis and Miller (1992) also asked listeners to rate the goodness of each stimulus using a ten-point scale ranging from 1 as the worst sounding /p/ to 10 as the best /p/. If stored categories do not contain rate-varied phonetic detail, the results would demonstrate no difference between ratings of short and long VOT stimuli as a function of rate. The authors found that listeners' goodness judgments were also subject to rate changes. VOT values of the highest rated /p/ were longer when syllable duration was longer. These results suggest that stored phonetic categories have detailed phonetic information encoding contextual speaking rate. Using a method of goodness judgment test, similar to the studies of Miller and Volaitis, Newman (2003) recently examined the correlation between speech perception and production within individual speakers. She found that the talkers who speak at a slower rate seem to have a perceptual prototype of /p/ with longer VOT values.

TABLE I. Mean and range of production VOT values (in ms) in the word initial /p/ and /b/ in English. Note: VDur and SDur stand for mean vowel duration and syllable duration in ms.

Source	Task	Speech rate	Mean VOT (Range)		VOT at boundary
			/b/	/p/	
Fast rates					
Miller et al. (1986)	Magnitude[a]	Fast (SDur=100–299)	⋯	⋯	14–27
Volaitis and Miller (1992)	Magnitude[a]	Fast (SDur=100–299)	9.6	45.2	⋯
Kessinger and Blumstein (1997)	Sentence[b]	Fast	13 (0–39)	63 (20–119)	20–29
Kessinger and Blumstein (1998)	Sentence[b]	Fast (VDur=60)	⋯	66 (30–110)	⋯
Allen and Miller (1999)	Isolated words[c]	Fast (VDur=149)	9		⋯
Allen and Miller (1999)	Isolated words[c]	Fast (VDur=141)		39	⋯
Normal/Medium rates					
Lisker and Abramson (1964)	Sentence[b]	Normal	7 (0–15)	28 (10–45)	⋯
Lisker and Abramson (1964)	Isolated words[c]	Normal	1 (0–5)	58 (20–120)	⋯
Miller et al. (1986)	Magnitude[a]	Medium	⋯	⋯	24–47
Volaitis and Miller (1992)	Magnitude[a]	Medium (SDur=300–499)	13.1	80.2	⋯
Unspecified rates					
Kessinger and Blumstein (1997)	Isolated words[c]	NA	13	88	20–39
Newman (2003)	Imitation[d]	NA	⋯	73 (51–125)	⋯
Slow rates					
Miller et al. (1986)	Magnitude[a]	Slow (SDur=500–799)	⋯	⋯	29–63
Volaitis and Miller (1992)	Magnitude[a]	Slow (SDur=500–799)	13.8	103.1	⋯
Kessinger and Blumstein (1997)	Sentence[b]	Slow	15 (0–39)	95 (40–149)	30–39
Kessinger and Blumstein (1998)	Sentence[b]	Slow (VDur=95)	⋯	91 (70–130)	⋯
Allen and Miller (1999)	Isolated words[c]	Slow (VDur=201)	6		
Allen and Miller (1999)	Isolated words[c]	Slow (VDur=173)		61	⋯

[a]Syllables elicited by a magnitude production technique.
[b]Words produced in a sentence.
[c]Words in an isolated condition.
[d]Nonsense syllables in an isolated condition. The model stimuli were presented before each speaker produced the test syllables.

Table II summarizes prior perception studies of rate effects on voicing perception. Only results obtained for English initial bilabial stops are included in Table II to make a direct comparison with the current study. It reveals a striking common factor among these studies. Stimuli used in these perception experiments were either synthetic or edited speech. When edited speech was used, VOT was increased or decreased by adding or removing a portion of syllable onset, or cross-splicing the /p/-onset with the /b/-onset and vice versa. Although previous studies support the claim that the internal representation of phonetic categories contains detailed subcategorical information, much of the research into rate normalization employs synthetic or artificially manipulated speech. As far as we know, there is no perception study on rate effects conducted using naturally produced rate-varied speech. It is surprising that the characteristics of voicing categories examining VOT variation have been examined only with synthesized or modified speech, with sounds created in a manner that listeners have never perceptually encountered. While using synthetic or edited speech has the advantage of allowing researchers to manipulate only one variable (e.g., VOT) to create equally distributed samples ranging from one category to another, there are potential problems raised by using synthetic or manipulated speech as we will discuss further below.

TABLE II. VOT values at the perceptual boundary (PB) for English initial bilabial stops. Note: The duration of stimuli used in Lisker and Abramson (1964) and Abramson and Lisker (1985) are not available, but grouped into the studies used short stimuli based on the short VOT at the perceptual boundary. Only the duration of VOT are available in Repp and Linn (1991) and Newman (2003) Most of the studies listed below varied VOT duration and the fixed duration for the rest of syllable except Miller and Volaitis (1989), Volaitis and Miller (1992), Miller *et al.* (1984a), and Pisoni and Tash (1974). Newman and Sawusch (1996) varied the duration of adjacent phonemes /l/ and /o/ separately, which are indicated as "varying /l/ dur" and "varying /o/ dur" in the table.

Source	Task	Stimuli type	Stimuli duration	VOT at PB
		Short (fast rate) stimuli		
Lisker and Abramson (1970)	Consonant ID	Synthetic speech[a]	···	About 23 ms
Abramson and Lisker (1985)	Consonant ID	Synthetic speech[b]	···	Around 15–25 ms
Summerfield (1981)	Consonant ID	Synthetic speech[c]	95–140 ms	28.1 ms
Miller and Vollaitis (1989)	Consonant ID	Synthetic speech[c]	125 ms	35.61 ms
Vollaitis and Miller (1992)	Consonant ID	Synthetic speech[c]	125 ms	39.18 ms
Repp and Lin (1991) [data of Exp 3 in Repp and Lin (1989)]	Consonant ID for "bin"-"pin" series (without precursor)	Edited speech by splicing	VOT (0–61 ms)+ /ɪn/	35.2 ms
Repp and Lin (1991)	Consonant ID for "bin"-"pin" series (without precursor)	Edited speech by splicing	VOT (10–53 ms)+ /ɪn/	23–52 ms
Summerfield (1981)	Consonant ID in Exp. 6	Synthetic speech[c]	167.5–212.5 ms	23.0 ms
Summerfield (1981)	Consonant ID in Exp. 6	Synthetic speech[c]	170–215 ms	29.5 ms
Miller *et al.* (1984)	Consonant ID for "beef"-"peef" series	Edited speech by splicing	210 ms	About 20 ms
Newman and Sawusch (1996)	Goodness rating with a 6-point scale for /blos/-/plos/ short series in Exp. 5	Edited speech by splicing	187–224 ms (varying /l/) 136–173 ms (varying /o/)	34.9 ms 29.6 ms
Newman and Sawusch (1996)	Goodness rating with a 6-point scale for /blos/-/plos/ long series in Exp. 5	Edited speech by splicing	239–276 ms (varying /l/ dur)	38.4 ms
Newman (2003)	Goodness rating with a 10-point scale in Exp. 1	Edited speech by splicing	8.25–291 ms (=VOT)+/ɑ/	44.7 ms[d]
		Long (slower rate) stimuli		
Pisoni and Tash (1974)	Consonant ID	Synthetic speech[a]	300 ms	About 30 ms
Newman and Sawusch (1996)	Goodness rating with a 6-point scale for /blos/-/plos/ long series in Exp. 5	Edited speech by splicing	299–336 ms (varying /o/ dur)	39.6 ms
Miller and Vollaitis (1989)	Consonant ID	Synthetic speech[c]	325 ms	43.89 ms
Vollaitis and Miller (1992)	Consonant ID	Synthetic speech[c]	325 ms	45.22 ms
Miller, *et al.* (1984b)	Consonant ID for "bath"-"path" series embedded in a sentence	Edited speech by splicing	425–491 ms	34.8–36.5 ms
Miller *et al.* (1984a)	Consonant ID for "beef"-"peef" series	Edited speech by splicing	431 ms	About 27 ms

[a]Synthesis by a parallel resonance synthesizer.
[b]Synthesis by the Haskins Lab formant synthsizer. The version of synthesizer was not reported.
[c]Synthesis by a serial resonance synthesizer.
[d]This is not equivalent to the /p/-/b/ boundary because the value was taken from the VOT at the lowest end of the VOT continuum for the /p/ category.

We can see another commonality in these studies. Rate effects on perceptual judgments were explored with a small number of rate conditions. Furthermore, in some of the perception studies (e.g., Miller and Volaitis, 1989; Volaitis and Miller, 1992), stimuli were blocked by syllable duration when presented to the listeners. In these cases, if syllable duration is defined as speech rate, listeners did not need to calibrate rate of speech during each trial. In order to understand the on-line perceptual normalization mechanism, we need to explore how listeners perceive speech sounds at various rates.

Finally, comparing Tables I and II reveals a systematic inconsistency between the production studies and the perception studies. Table I shows previously found mean and/or range of produced VOT values[1,2] for initial bilabial stops and estimated /b/-/p/ boundaries, which are listed in terms of

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

K. Nagao and K. J. de Jong: Perceptual rate normalization   2885
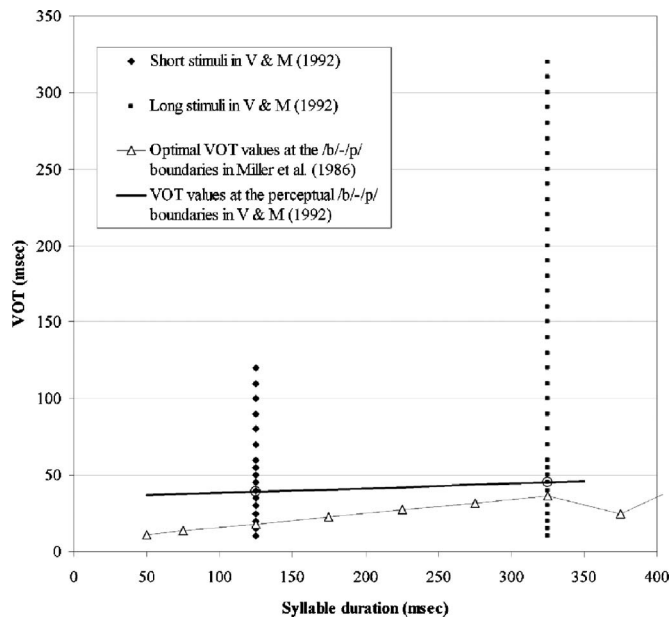
FIG. 1. The filled diamonds and the filled squares represent the short and long stimulus used in the perception experiments in Volaitis and Miller (1992). The unfilled triangles interpolated with a thin line represent the rate-dependent optimal VOT to yield 90% or higher categorization performance for /bi/ and /pi/ productions within successive 50-ms syllable duration interval reported in Miller et al. (1986). The thick line represents VOT at the perceptual /b/-/p/ category boundaries reported in Volaitis and Miller (1992).

speech rate. Again, Table II lists the studies in terms of speech rate or stimuli duration. While the VOT values at the perceptual boundary become longer with longer stimuli durations, the actual reported values are not consistent with the production boundaries. The stimuli with 200 ms or shorter duration, i.e., fast rate stimuli, showed the perceptual boundaries at between 15 and 52 ms of VOT. The VOT at the production boundaries do not exceed 30 ms, as seen in Table I. For the stimuli with syllable duration between 300 and 500 ms, the VOT at the perceptual boundary appeared between around 27 and 45 ms, which is consistent with the production boundaries for medium rate speech, but exceeds the VOT production boundaries reported for slow speech rate.

Specifically comparing the results by Miller and her colleagues, we find that the estimated perceptual boundary [for example, reported in Miller and Volaitis (1989) and Volaitis and Miller (1992)] do not match with the estimated production boundaries in Miller et al. (1986). Figure 1 superimposes the /b/-/p/ category boundary estimated in the production study (Miller et al. 1986) and the estimated boundary in the perception study (Volaitis and Miller, 1992) in the stimuli space employed in Volaitis and Miller (1992). Volaitis and Miller (1992) employed the two series of VOT continua. Filled diamonds in Fig. 1 represent their short stimuli (i.e., fast stimuli) with the syllable duration of 125 ms, and the filled squares represent their long stimuli (i.e., slow stimuli) with the syllable duration of 325 ms. Unfilled triangles on the thin line represent the rate-dependent optimal VOT values estimated as category boundaries in Miller et al. (1986).

As is clearly seen in Fig. 1, perceptual VOT boundaries (thick line) were at much higher values than those estimated from the production studies (thin line).

Interestingly, a similar mismatch between production and perception has been found in a vowel study (Johnson et al., 1993). Johnson et al. (1993) compared speakers' vowel productions and their perceptual vowel space. In order to examine listeners' vowel spaces, a method of adjustment task was employed. Johnson et al. (1993) created synthetic vowels by changing the first and second formants (F1 and F2). A set of synthetic sounds for each vowel was arrayed in a two-dimensional grid of F1 and F2 on a CRT screen so that listeners could hear the sound associated with each cell in the grid. Prior to the perception experiment, subjects were asked to produce target vowels embedded in a word. Subjects were then asked to click a cell in the grid to locate the sounds that match a target vowel displayed on the same screen. The perceptual responses were systematically more extreme in formant pattern than their productions. Based on the result that the perceptual vowel space is expanded similar to the vowel space in hyperarticulated vowels, Johnson et al. (1993) called this effect the hyperspace effect. Although the methodology and purpose of their study are quite different from the studies of Miller and Volaitis, the mismatch between production and perception is strikingly similar.

More recently, the interpretation of Johnson et al.'s results as a hyperspace effect has been questioned (Whalen et al., 2004a, b; but see Johnson et al., 2004). The crux of the criticism is that there is really no way of determining what the actual production space for a synthetic matrix of vowels is. Hence, perceptual responses could be a reflection of listeners' expectations about indexical differences between the perceiver and their estimate of who (or what) might have produced the matrix of vowels. Whalen et al. (2004b, c) replicated the study of Johnson et al. (1993) with speakers of a different dialect (Rhode Island) and found that back vowels showed more frontness (hypospace effect) rather than more backness (hyperspace effect) in perceptual vowel space.

However, a hyperspace effect has also been reported for English consonants using a somewhat different paradigm (Newman, 2003). Newman (2003) examined the correlations between speech production and perception for consonants within individuals. She examined the correlation between the average VOT value for /p/ that individual speakers produced and the VOT value of /p/ stimulus that each person rated as the best member of the category. The stimuli used in her experiment 1 were created by modifying VOT values of natural speech to represent /b/ to /p/ to exaggerated *‍/p/. A majority of the subjects (75%) showed the result that the highest rated /p/ stimuli had longer VOT values than their own average productions.

Comparing these studies suggests four possible reasons for the mismatch in production and perception in the rate normalization literature. First would be some sort of direct hyperspace effect induced by using synthetic or edited stimuli. Since synthetic speech and targeted editing generally fail to capture the many covarying aspects of the signal that indicate the voicing category, listeners may "load up" on the VOT dimension to make up for the missing secondary cues.

If we assume the best exemplars correspond to the hyperarticulated phonetic targets, the best exemplars should have longer VOT for /p/ and shorter VOT for /b/ than the VOT values typically found in speech. If both categories have hyperarticulated phonetic targets that are expanded symmetrically from a VOT category boundary, the hyperspace effect would not affect the location of the perceptual boundary. However, production studies show that the /p/ category varies more than the /b/ category, and hence the /p/ category is likely to have a larger hyperarticulated shift.

A second reason could be the choice of stimulus space with respect to what speakers actually produce. Miller and Volaitis (1989) and Volaitis and Miller (1992) employed quite extreme stimuli [as did Johnson *et al.* (1993) and Newman (2003)]. Figure 1 shows how wide a range of VOT was employed in Volaitis and Miller (1992) (the series of dots to the left and right). Stimuli with extremely long VOT were used as representations of an exaggerated */p/. Hence the stimulus set for /p/ might be expanded to include this extreme category */p/ as a subcategory of /p/. If this is so, given the stimuli with a wide range of VOT as well as an expanded /p/ category, listeners' phonetic targets might be altered for the specific task in the identification protocol and cause a shift in the perceptual boundary toward longer VOT. Then, it is possible that the mismatch between production and perception boundaries was due to a task space effect introduced by using such extreme stimuli.

Alternatively, the inclusion of these extreme stimuli could also create a perception-production mismatch because of the introduction of a three-category system that differs from the naturally produced two-category system of English. Subjects in Miller and Volaitis (1989) and Volaitis and Miller (1992) listened to all the stimuli once in order of increasing VOT (e.g., from /bi/ to /pi/ to */pi/). They were also told to pay attention to where the sound starts to change from /bi/ to /pi/ and from /pi/ to */pi/. As pointed out by Utman (1998), this process of familiarization might impose an external (three-category) system on the listeners so that the results might not actually reflect the structure of the listeners' normal categories. Utman (1998) found that the familiarization process with an additional category of voiceless sounds does affect perceptual judgments on the normal voiceless sounds. As Utman (1998) noted, "the use of a third category in the identification task may have artificially narrowed the best-exemplar range for voiceless consonants by causing subjects to interpret these stimuli as falling into one of two categories" (p. 1647).

Finally, the perception-production mismatch might also be due to the way the stimuli were created. Unless a precursor phrase was given, speaking rates were defined only in terms of syllable duration in previous studies. In order to create VOT continua, syllable durations were fixed at either a short syllable duration or a long syllable duration, while VOT was modified in proportion to the total syllable duration. However, VOT is not the only change accompanying rate change. When speaking rate decreases, both vowel duration and VOT become longer (Allen and Miller, 1999; Crystal and House, 1988; Peterson and Lehiste, 1960). Also, vowels following voiceless stops are shorter than vowels fol-

lowing voiced stops, and speakers maintain the difference in vowel duration between voiceless and voiced stops (Allen and Miller, 1999). Synthetic speech and artificial editing do not perfectly reflect the complexity of temporal dynamics within a syllable; the relation between artificial rate changes and the natural temporal variation that listeners are accustomed to is unclear, and hence the effect of artificial modulation on perception expected on the basis of the listeners' experience can often be unclear. For example, Kessinger and Blumstein (1998) pointed out that VOT continua with fixed syllable duration could alter perception of the vowel following the initial consonant from /i/ to /ɪ/ because the artificial vowel durations are not typical of the intended vowel /i/ (cf. Allen and Miller, 1999).

The purpose of the current study was to examine speech rate effects on the perception of voicing contrasts in natural speech. Using natural speech eliminates most of the difficulties mentioned above. Natural speech reflects the dynamics of the production processes and includes the general distribution of temporal variation that listeners are familiar with. Stimuli here were generated with a rate control paradigm to vary across an extreme range of rates found to be practical by previous speech research. Experiment I was conducted to examine the rate effects on productions of syllables /bi/ and /pi/. Experiment II explored the rate effects on perceptual identification of these rate controlled productions. The boundary estimates from Experiment I and the perceptual boundary estimates from Experiment II were compared with one another and with those of previous studies to determine if using naturally produced stimuli eliminates the production-perception mismatch found in previous studies. Experiment III was conducted to examine the effect of rate on listeners' estimates of best exemplars, to determine if a hyperspace effect is present in the perception of naturally produced stop consonants.

II. EXPERIMENT I: ACOUSTIC ANALYSIS

Experiment I examined the speech rate effects on the VOT distributions of /b/ and /p/. Although acoustic analyses have been done for speech produced at various rates (Miller *et al.*, 1986; Volaitis and Miller, 1992), rates were not externally controlled but manipulated by speakers. Miller and her colleagues collected syllables at various rates with a magnitude production procedure. In this method, speakers were asked to repeat each syllable six times first at normal rate, and then to produce two slower and two faster rates relative to this normal rate, and in addition a slowest and a fastest possible rate. In Miller *et al.* (1986), the syllable duration of their samples ranged from 111 to 700 ms. Although this range covers the average syllable durations observed in spontaneous speech, their samples can be thought of as including a lot of slow speech, when compared with reported average syllable durations in American English, which ranges from 100 to 300 ms in spontaneous speech (Goldman-Eisler, 1968; Miller *et al.*, 1984c; Roach, 1998). It is important to examine the rate effect with syllables produced at various rates, especially at fast rates, since self-controlled rates in laboratory speech tend to be slower than usual. The aim of

experiment I was to observe the VOT values for /b/ and /p/ productions at various rates systematically sampled, especially at fast rates, and to examine how the VOT values used for the /b/-/p/ boundary change as syllable duration increases.

A. Methods

1. Subjects

Two male and two female speakers participated in the recording at the Indiana University Phonetics Laboratory. All were native speakers of Midwestern varieties of American English and were in their mid-30s at the time of recording. All had no history of any speech and language disorders.

2. Speech materials

Stimuli were taken from a speech corpus originally collected for other purposes. The corpus includes productions of four native speakers of American English (two male, two female). In this corpus (de Jong, 2001a, b), speakers were asked to repeat the syllables /bi/, /pi/, /ib/, and /ip/ in time with a metronome. They repeated the same syllable for approximately 20 times with increasing, decreasing, or fixed rates of repetition. They were presented with a pulse train that began with a series of six clicks with a period of 400 ms. After the six clicks, the period was decremented by 12.5 ms per period until reaching a period of 150 ms. The train concluded with a series of six more clicks at the fixed rate of 150 ms. Before the set was begun, the speakers were warned that the pacer would change rate. This elicitation technique has proved to be an easy task for naive subjects. Only the CV syllables (/pi/ and /bi/) from the increasing rate condition were used for the current study. The fastest 21 repetitions of each syllable were subject to the acoustical analysis. In order to compare our results with previous studies, open syllables whose onset consonant was a labial stop (/bi/ and /pi/) were used for acoustical analysis. The total number of tokens was 168 (=4 speakers by 2 consonants (/bi/ and /pi/) by 21 rates).

3. Measurements

For each syllable, the beginning of the stop release, the beginning of the modal voicing for the vowel, and the end of the vowel were determined by a visual inspection of the waveforms and spectrograms [for more information on measurements, see de Jong (2001a, b)]. Following Miller *et al.* (1986), the value of the VOT interval was defined as the time interval from the consonant release to the beginning of the voicing, and syllable duration as the duration from the release to the end of the vowel. Note that these measurement criteria make our definition of VOT somewhat different from that sometimes used in the literature, where stops with prevoicing have negative VOT values. In previous work with this corpus of productions, the existence of prevoicing is considered as a separable acoustic attribute from VOT, which measures degree of gap between the burst and modal voicing. Systematically, the current stimuli have no voicing in the stop closure at slow rates, and both /pi/ and /bi/ stimuli gain it at faster rates (de Jong, 2001a).
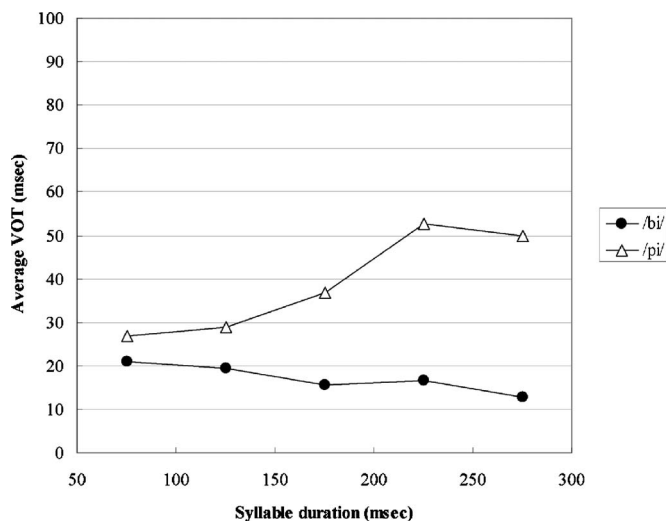
FIG. 2. Mean VOT values of /bi/ and /pi/ tokens within successive 50-ms intervals of syllable duration. The filled circles represent /bi/ tokens and the unfilled triangles /pi/ tokens.

In order to assess measurement reliability, six stimuli (three /bi/ stimuli and three /pi/ stimuli) from each of the four talkers were selected at random, and values of VOT and syllable duration for these 24 stimuli were measured by a different phonetician (the first author). Pearson correlation coefficients were computed for the two measurements for each stimuli type. The results of high r values suggest that measurement reliability was adequate for this study (VOT for /b/ stimuli: $r=0.91$; VOT for /p/ stimuli: $r=0.96$; syllable duration for /b/ stimuli: $r=0.95$; syllable duration for /p/ stimuli: $r=0.96$). Average differences between the two measurements were small for both /b/ and /p/ stimuli (VOT: $M=1.1$ ms for /b/ stimuli and $M=1.5$ ms for /p/ stimuli; syllable duration: $M=3.6$ ms for /b/ stimuli and $M=2.7$ ms for /p/ stimuli).

B. Results

In order to see the relationship between VOT and syllable duration, all the syllables except one[3] were grouped into 50-ms bins in terms of the syllable durations, and the VOT values of the four speakers were averaged for each bin of syllable duration. Figure 2 plots average VOT values of /bi/ and /pi/ as a function of syllable duration. The relationship between VOT and syllable duration for /pi/ is similar to that reported in previous literature (Miller *et al.*, 1986; Miller and Volaitis, 1989; Volaitis and Miller, 1992; Kessinger and Blumstein, 1997). VOT increases as syllable duration increases. As for /bi/, VOT does not change much as a function of syllable duration. Although the differences are small, the slope function for /bi/ is slightly negative, which means that VOT actually decreases as syllable duration increases.

In order to illustrate the detailed distributions of VOT values for /bi/ and /pi/, the VOT values of all the tokens produced by four talkers were plotted against their syllable durations in Fig. 3. In Fig. 3, a positive linear relationship between VOT and syllable duration is displayed for /p/ productions. Although the VOT values for /p/ increase as syllable duration increases, the VOT values for /b/ do not

K. Nagao and K. J. de Jong: Perceptual rate normalization
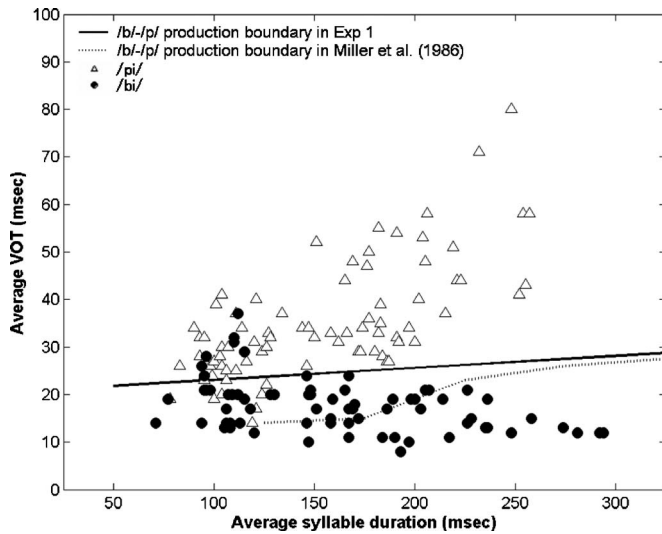
FIG. 3. VOT values for /bi/ and /pi/ tokens plotted as a function of syllable duration (/pi/: $r=0.68$, /bi/: $r=-0.40$). The filled circles represent /bi/ tokens and the unfilled triangles /pi/ tokens. The straight line represents estimated /b/-/p/ boundary locations, and the dotted line the /b/-/p/ boundary locations reported in Miller et al. (1986).

change much, as seen in Fig. 3, and, if anything, tend to shorten at slower rates. The actual slope values were 0.1685 for /p/ and −0.0413 for /b/. The VOT values of each consonant are well separated when syllables are relatively long (>200 ms); however, they overlap considerably with each other when the syllables are short.

A binary logistic regression analysis (e.g., Menard, 2001) was performed with VOT and syllable duration (syllable) as predictor variables. The response variable was defined as whether the speakers' intended syllables were either /p/ (=1) or /b/ (=0). The probability of /p/ was computed by the equation

$$\rho(|p|) = \frac{e^{\beta}}{1 + e^{\beta}} \qquad (1)$$

where e is the natural logarithmic base and β is given by the following equation:

$$\beta = -4.564 + 0.232(\text{VOT}) - 0.007(\text{syllable}). \qquad (2)$$

Both VOT and syllable variables were significant predictors of the speakers' intended consonant (p's <0.001). The regression equation correctly classified 86.9% of /b/ tokens and 88.1% of /p/ tokens for a total predictive efficiency of 87.5%. The category boundary was estimated from VOT at the 50% point of the regression function; it is plotted as a solid line in Fig. 3. It is clear that the optimal criterion VOT is rate sensitive, with a positive slope as a function of syllable duration.

Miller et al. (1986) estimated the rate-dependent optimal /b/-/p/ boundaries by locating the VOT value to yield best categorization performance (greater than 90% for their data) for the data sets within successive 50-ms syllable duration between 100 and 700 ms. Their estimated VOT values for the /b/-/p/ boundaries are plotted as a dotted line in Fig. 3. We can see both boundary lines are close to each other at slower rates (to the right). Their estimated optimal bound-

aries successfully separate our speech samples into /b/ and /p/ categories at slower rates, but errors start to occur at faster rates. Their estimated boundaries appear in the middle of /b/ productions when the syllable has 200 ms or shorter duration. Comparing our boundaries to those in Table I shows that our results are quite similar to previous results, and the estimated boundary actually is very close to that found by Kessinger and Blumstein (1997). Although our estimated /b/-/p/ boundary line is not as steep as the boundary reported in Miller et al. (1986), it is nevertheless clear that VOT at the category boundary increases with increasing syllable duration.

III. EXPERIMENT II: PERCEPTION EXPERIMENT I

Experiment II was conducted to examine speech rate effects on voicing identification. The aims were to examine whether the category boundary shift observed for synthesized and edited sounds occurs in the same way when listeners judge natural speech or not, or whether the perceptual category boundary matches the category boundary estimated from the production studies. We prepared the stimuli from the same speech corpus used in experiment I. Our speech samples include faster rates of speech than the data set of Miller et al. (1986). In their studies, syllables with less than 100 ms duration were not observed. However, syllable durations of our speech samples ranged from 68 to 331 ms. All the syllables except one used in experiment II were under 300 ms in duration. This matches with average syllable durations in American English, but could be considered in the range of the fast rate in the study by Miller et al. (1986).

There were also several methodological differences between the current study and the studies by Miller and Volaitis (1989) and Volaitis and Miller (1992). Miller and Volaitis (1989) and Volaitis and Miller (1992) employed three choices (/b/, /p/, and *_/p/) for the identification task. We assumed that the use of unnatural category *_/p/ was not necessary since the stimuli in experiment II were all naturally produced /b/ or /p/, and *_/p/ was not an intentional category of the speakers. Also, it is not necessary to include a process of familiarization with the stimuli. In Miller and Volaitis (1989) and Volaitis and Miller (1992), stimuli with the same syllable duration were blocked, so listeners did not need to constantly normalize speech rate for every trial. The stimuli in experiment II had gradiently varied syllable durations, and they were randomly presented to the listeners to see the effects of local rate on voicing categorization. Listeners were asked to listen to a portion of repetitive speech that we collected in experiment I and identify the consonant that the speaker intended to produce.

The category boundary between /b/ and /p/ was estimated in terms of VOT and syllable duration based on the listeners' identification in order to examine the consistency of production and perception. We compared the /b/-/p/ category boundary estimated from the current study with the one obtained for synthesized sounds (Miller and Volaitis, 1989; Volaitis and Miller, 1992) as well as with the one estimated in experiment 1. When the listeners judge natural sounds, if the boundary shift occurs in the same way as in
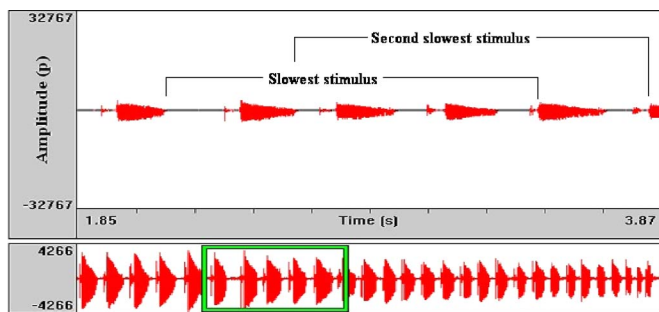
FIG. 4. (Color online) Stimuli used in the perception experiments.

previous studies, we can generalize the results of artificially manipulated or synthesized sounds to natural sounds. Unlike synthesized speech, natural speech preserves all the acoustic information during the transition from a consonant throughout the following vowel, so any relevant acoustic property can be used for voicing categorization when the rate changes. In addition, we presented the listeners repeated syllables as a stimulus in order to provide the listeners with rate information.

A. Method

1. Subjects

Twenty-three native listeners of American English from 18 to 24 years old participated as paid subjects in the study. Data from one subject were not included, because a hearing problem was self-reported. Four bilingual subjects were excluded from further analysis because the voicing distinction in the other language might influence their perception of the voicing contrast in English. Data from 18 native listeners of American English were subject to further analysis.

2. Stimuli

Speech materials used for the stimuli in experiment II were the same as used in experiment I. In the production corpus from which the stimuli were drawn, the four speakers repeated the same syllable (/bi/, /pi/, /ib/, and /ip/) with increasing rate of speech. All four of these types were submitted to perceptual identification procedures; however, since previous studies of rate normalization focused on prevocalic stops, the current study only focuses on results for the prevocalic (/bi/ and /pi/) productions.

From each repetitive utterance, 21 stimuli were spliced. Each of these 21 stimuli consists of three repeated vowels. Three syllable nuclei were included in each stimulus in order to allow for perception of both post- and prevocalic consonants in the middle syllable. Splicing any of the abutting syllables out would inevitably remove information about the identity of the consonant. Analyses of perceptual results using different splicing locations, with and without consonant transients on both ends of the three-vowel stimuli, show that giving listeners a train of three syllables gives them enough context to allow them to abstract away from the edges created by the splicing technique and give consistent identifications across splice locations (de Jong *et al.*, 2004). Figure 4 displays waveforms of a repetitive utterance of /pi/ by a female speaker. The slowest stimulus was spliced from the

vowel offset of the fifth slowest syllable to the vowel onset of the ninth slowest syllable, and the second slowest stimulus was spliced out from the vowel offset of the sixth slowest syllable (or the first syllable of the slowest stimulus) to the tenth slowest syllable in the same utterance. In this way, 21 stimuli were prepared from four types of repetitive utterances by four speakers. The total number of stimuli was 336. The current paper analyzed the responses for 168 out of 336 stimuli because responses for the stimuli /ip/ and /ib/ were excluded for further analysis.

B. Procedure

Listeners were instructed that they were to listen to a portion of speech whose speaker was originally repeating the same syllable over and over. They were asked to listen to each sample carefully and to answer what they think the speaker was repeating from the four choices ("pea," "bee," "eep," and "eeb"). Presentation of the stimuli was controlled by a customized Matlab protocol. Listeners listened to the stimuli through Sony MDR-CD280 Stereo Headphones in a computer laboratory and selected one of the four choices on a computer screen by clicking a mouse. They were allowed to listen to each stimulus as many times as they wished. Each listener was given each of the stimuli once in a different random order. Prior to the experiment, the listeners had eight trials as practice. Eight stimuli used in the practice session were not used in the experiment session. During the experiment, a short break was given every 100 trials. Most of the listeners completed the whole experiment in less than one hour.

C. Results

Responses for the 15 stimuli identified as either /ip/ or /ib/ among more than 75% of the listeners were excluded for further analysis. Twelve out of the 15 excluded stimuli were /bi/ stimuli. All excluded stimuli were fast rate stimuli (i.e., average syllable duration was less than 125 ms) except one stimulus (the average syllable duration was 137 ms). Hence, the responses for the 153 stimuli from 18 listeners were analyzed.

Figure 5 plots VOT values for /bi/ responses (filled circles) and /pi/ responses (unfilled triangles) against syllable duration. Both VOT and syllable durations were taken from the averaged duration of the second and the third syllables in each stimulus. Stimuli with the average syllable durations of 200 ms or longer will be categorized as slow rates, stimuli with 125 ms or shorter average syllable duration will be categorized as fast rates, and the rest of the stimuli as mid rates. The size of markers indicates the performance levels on voicing perception; the larger the marker, the more listeners correctly identified the consonant. In general, listeners' consonant voicing identification was accurate for both /bi/ and /pi/ stimuli. There were 12.8% incorrect responses among the total of 2754 responses. Most of the errors (89%) occurred for the /bi/ stimuli, and most of these occur at fast rates. The /bi/ stimuli with 125 ms or shorter average syllable duration were responsible for 74.9% of the incorrect responses. Thus, while listeners identified both consonants very accurately at
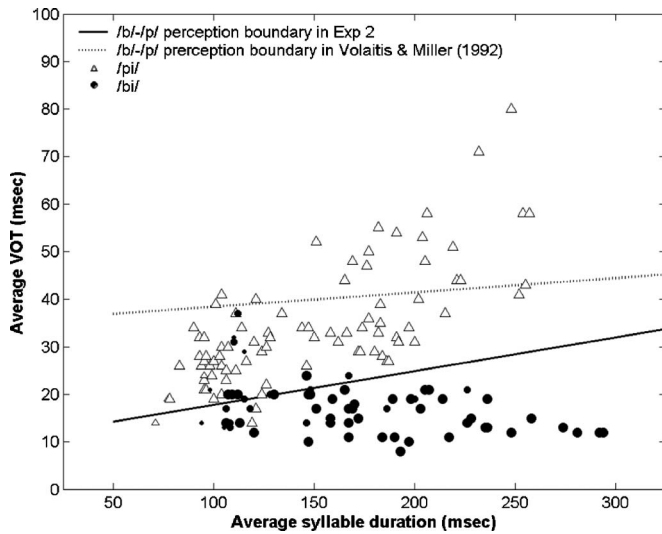
FIG. 5. VOT values for /bi/ and /pi/ responses plotted as a function of syllable duration (/pi/: $r=0.71$, /bi/: $r=-0.36$). The filled circles represent /bi/ percepts and the unfilled triangles /pi/ percepts. The size of the markers indicates mean percent correct identifications of the consonant by 18 listeners in experiment 1. (The size decreases as mean percent identifications decreases.) The straight line represents estimated /b/-/p/ boundary locations, and the dotted line the /b/-/p/ boundary locations reported in Volaitis and Miller (1992).

slow rates (stimuli with 200 ms or longer average syllable duration), performance of /b/ identification decreases at fast rates and performance of /p/ identification remained high. In other words, listeners tend to perceive the consonant in fast rate syllables as /p/.

In order to estimate the perceptual boundary, a binary logistic regression analysis was performed with average VOT and syllable duration of the second and the third syllables in each stimulus as predictor variables. The response variable was whether the listener responded to the stimulus as /p/ (=1) or as /b/ (=0). The probability of /p/ perception was computed by Eq. (1). The equation for the relationship between the dependent variable and the independent variables was

$$\beta = -2.394 + 0.225(\text{VOT}) - 0.016(\text{syllable}).\qquad(3)$$

Both VOT and syllable duration contributed significantly to the prediction of consonant identification [p's < 0.001]. The logistic regression equation accounted for 85.3% of the dependent variable's variance. The percentage of correctly classified /pi/ stimuli was 88.0% and was greater than the correct classification of /bi/ stimuli (80.7%).

Analysis based on the VOT and syllable durations of the middle syllable in each stimulus was also conducted. The results turned out to be similar to the analysis based on averaged durations. The equation was

$$\beta = -1.441 + 0.194(\text{VOT}) - 0.017(\text{syllable}).\qquad(4)$$

In the analysis based on the second syllable durations, the percentage of correctly classified stimuli slightly decreases to 87.2% and 79.2% for /pi/ and /bi/ stimuli, respectively.

VOT values at the perceptual /p/-/b/ boundary were estimated by calculating the VOT values to yield 50% /p/ responses for syllable durations between 50 and 325 ms. The

estimated boundary is shown in Fig. 5 (solid line). The positive slope of the estimated boundary function indicates a strong rate dependence such that as syllable duration increases, boundary VOT values increase. Figure 5 also displays the perceptual category boundary between /bi/ and /pi/ reported by Volaitis and Miller (1992) (dashed line). It is clear that the perceptual boundary obtained with the current data has considerably shorter VOTs than that obtained from the listeners' responses in the study by Volaitis and Miller (1992). Comparing the boundary obtained in experiment II to those generally found with synthesized or edited speech summarized in Table II shows that, while not all previous studies obtain such high VOT values as Volaitis and Miller (1992), all of these previous studies [with the possible exception of Summerfield (1981) and Miller *et al.* (1984a)] obtain higher VOT boundaries than are obtained in experiment II.

Comparing Fig. 3 to Fig. 5 shows that the production boundary, while somewhat greater in VOT than our current perceptual results, is still closer to the perceptual boundary found in experiment II than to those found in previous studies. Examining the pattern for individual tokens shows that the production and perception boundary is remarkably close for syllable durations greater than 100 ms. The only systematic deviation obtained here is for the very short syllable durations where a majority of /b/ responses are virtually nonexistent. The unidirectional misperception result that /p/ was not perceived as /b/ at very fast rate seems to also support the claim that the hyperarticulated targets for /b/ and /p/ in English are asymmetrically distributed. This issue will be more directly addressed in experiment III.

To summarize, speech rate affects voicing perception in naturally rate-varied speech. This result is consistent with previous findings using synthesized and artificially edited sounds. However, the estimated perceptual boundary is closely matched with the category boundary estimated from productions in experiment I. Deviations of the production and perception functions are mostly due to misidentification of /b/ stimuli with very short syllable durations. Aside from this discrepancy, the current results suggest that previous production-perception mismatches may be due to the use of synthetic or edited stimuli with unnatural distributions over the space of speech rates. Experiment III pursues the particular mechanism responsible for the mismatch by examining listeners' estimates of best exemplars to determine if a hyperspace effect can be detected with the current stimuli.

IV. EXPERIMENT III: GOODNESS RATING

The results of experiments I and II suggest that the perception-production mismatch found in the previous studies could be due to the use of artificial speech. In addition, the extreme VOT values employed in the previous studies could distort the distribution of each consonant category. Alternatively, if we assume the asymmetrical distribution of the hyperarticulated phonetic target for /b/ and /p/, the hyperspace effect should appear at a slow rate of speech even for the natural stimuli. The purpose of experiment III was to examine the rate effects on estimates of the best exemplars of

voicing categories for natural speech. Listeners were asked to rate the goodness of the consonant that they heard. We asked the listeners to identify the consonant and then judge the goodness of the consonant in terms of the consonant category they perceived.

A. Method

1. Subjects

Seventeen native listeners of English were newly recruited and participated in the study. They were undergraduate students of Indiana University whose ages were from 18 to 27 years old. All of them reported that they had not been diagnosed as having any hearing problems or language disorder. They were paid for their participation.

2. Stimuli

The same stimuli used in experiment II were used for this study. The 336 stimuli were presented to the listeners in random order, but the responses for the 168 coda stimuli (/ip/ and /ib/ stimuli) were excluded for further analysis, as we did in the analysis in experiments I and II.

B. Procedure

Listeners were given similar instructions as in experiment II: they were told that they were to listen to a portion of speech whose speaker was originally repeating the same syllable. There were two tasks in this study. One was to identify the consonant and the other one was to rate the goodness of the consonant. They were asked to listen to each sample carefully and to answer which consonant they thought the speaker was repeating from two choices (/p/ and /b/). This task differed from the four-way forced choice identification task in experiment II to insure that the difference between the current results and previous results was not due to a task that integrated syllable affiliation and consonant voicing in the same task. Then, listeners were asked to rate the goodness for the consonant they selected using a scale from 1 (= "Terrible") to 10 (= "Excellent"). Presentation of the stimuli was controlled by a customized Matlab protocol. As in experiment II, listeners were to listen to the stimuli through headphones and provide their answers by a mouse click. They were allowed to listen to each stimulus as many times as they wished. Prior to the experiment, eight practice trials were given to each listener. Listeners had a short break every 100 trials. The whole experiment usually was completed in less than one hour.

C. Results

The 15 stimuli excluded in experiment II were also excluded in the analysis in experiment III in order to compare the results of both perception experiments. Therefore the analysis of results was based on 2601 responses (=17 listeners by 153 stimuli). For the same reason, the values of VOT and syllable durations used for the analysis were the average values between the second and third syllables in each stimulus.
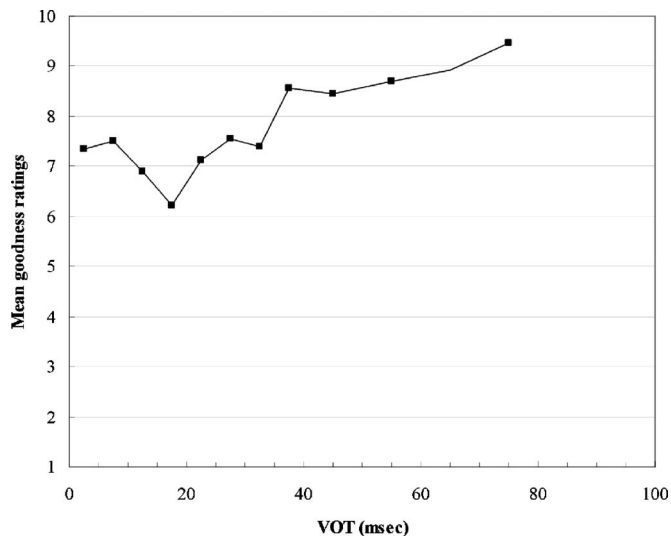
FIG. 6. Mean goodness ratings as a function of VOT for correctly identified /bi/- and /pi/-stimuli. Judgments were made on 10-point scales (1 = Terrible, 10 = Excellent).

The consonant identification results replicated the results of experiment II. Mean percent of /p/ responses for each stimulus in experiments II and III were separately arcsine transformed (Studebaker, 1985). A two-tailed paired t test of transformed mean percent /p/ responses revealed that there was no significant difference between experiment II and experiment III for the overall consonant identification results [$t(152)=0.59$, $p=0.56$]. Overall, listeners' consonant identification was accurate for both /bi/ and /pi/ stimuli. There were 9.8% incorrect responses among the total of 2601 responses. Among the incorrect responses, /b/ stimuli with syllable durations shorter than 125 ms were responsible for 83% of the incorrect responses. This result suggests that listeners tend to perceive the onset consonant in very fast syllables as voiceless. On the other hand, /p/ stimuli with 200 ms or longer syllable durations produced no misidentifications among the listeners. These results are virtually identical to those from experiment II, indicating that the difference between ours and previous results were not due to the integrated judgments on syllable and voicing in our task in experiment II.

Turning to the goodness judgments, Fig. 6 plots mean goodness ratings for the correct responses across all listeners against VOT. The VOT value and syllable duration for each stimulus are based on the average durations of the second and the third syllables. Responses were averaged for stimuli in 5 ms bins (for VOT <40 ms) or 10 ms bins. The lowest mean rating (=6.2) appeared for the stimuli with 15–20 ms VOT, which is the location of the category boundary. The highest mean ratings were found for the stimuli with the shortest VOTs and with the longest VOTs. Thus, the best exemplar of /p/ has an extremely long VOT, while the best exemplar of /b/ has an extremely short VOT. This general pattern is similar to the hyperspace effect found by Johnson *et al.* (1993), whereby best exemplars are far from category boundaries, and hence tend to be tokens that are extreme in phonetic attributes.

In order to see the relationship between rate and good-

TABLE III. Mean goodness ratings for correctly identified /b/ and /p/ stimuli in three rates (1 = "Terrible" to 10 = "Excellent").

Rate	/b/ stimuli	/p/ stimuli
Fast	5.41	6.67
Mid	7.20	8.00
Slow	7.67	8.82

ness rating, mean goodness ratings for correct responses were computed for three rates (fast, mid, and slow). Three rates were defined by average syllable durations of the second and the third syllable in each stimulus. Stimuli with average syllable durations of less than 125 ms were grouped into a fast rate category (fast), the stimuli with average syllable durations of longer than 200 ms were grouped into a slow rate category (slow), and the stimuli with average syllable durations of longer than 125 ms but less than 200 ms were grouped into mid rate (mid). Mean ratings are presented in Table III.

As is apparent in Table III, the consonants produced at slower rates received higher ratings. Also, the /p/ stimuli are more highly rated overall than the /b/ stimuli. A two-way ANOVA was performed on the mean goodness ratings across subjects for correct responses with rate (fast versus mid versus slow) and consonant (/b/ and /p/) as independent variables. Significant main effects were found for both rate and consonant [$F(2,147)=81.4$, $p<0.0001$; $F(1,147)=63.1$, $p<0.0001$, respectively]. The interaction effect was significant at $\alpha=0.05$ level, but not significant at 0.01 level ($p=0.02$). Thus for the stimuli in this study, regardless of voicing category, listeners judged that consonants produced at slow rates are better examples of each category. Although the reason is not clear, tokens in the voiceless category are con-

sistently judged with higher ratings than those in the voiced category, and this rating difference was especially large at fast rates.

To examine in more detail the effects of rate on the relationship between VOT and goodness rating, Fig. 7 separates /b/ stimuli (top panels) from /p/ stimuli (bottom panels) and gives separate functions for the three different rates (fast, mid, and slow). Figure 7 also includes separate functions for correctly identified stimuli in the left panels and for incorrectly identified stimuli in the right. In order to see the small differences in VOT for fast /b/ stimuli clearly, 2.5 ms bins were employed to group the stimuli with VOT values less than 20 ms. Among correctly identified /b/ syllables (top left panel), stimuli with shorter VOTs were rated higher, while stimuli with longer VOTs, which is closer to the boundary with /p/, received lower ratings. The one exception to this pattern is for the slowest rates (filled triangles), where the listeners preferred slightly longer VOTs. The ratings for /p/ syllables (bottom left panel) had the opposite trend, with stimuli with long VOTs generally receiving the higher ratings and stimuli with short VOTs closer to the boundary with /b/ receiving lower ratings (though for some reason the stimuli with the shortest VOT at fast rates and the mid-rate stimuli with 30–35 ms VOTs tended to be rated very high). The most frequent consonant misidentification occurred for the stimuli with 7.5–25 ms VOTs, which is again consistent with the critical VOT values that had been reported for the /b/-/p/ boundaries in previous literature (see Table II). On the whole, the listeners tended to rate /b/'s with relatively short VOTs and /p/'s with relatively long VOTs as better than their more moderate counterparts.

Furthermore, listeners gave higher ratings when they correctly identified the voicing of the consonant than when they misidentified it, except that the fast /b/ stimuli misper-
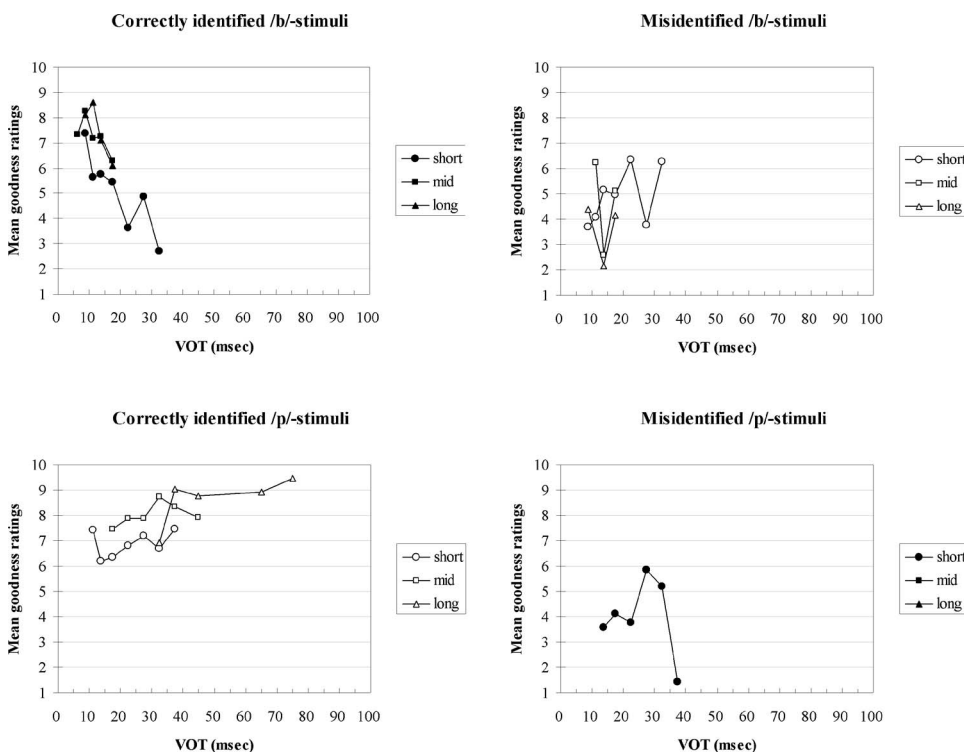
FIG. 7. Mean goodness ratings for correctly identified /bi/ stimuli (top left panel), /pi/ stimuli (bottom left panel), misidentified /pi/ stimuli (top right panel), and misidentified /bi/ stimuli (bottom right panel) as a function of VOT at the three rates (fast, mid, and slow). Three rates are defined by the average syllable durations (SDur) of the second and the third syllables in each stimulus: fast (SDur ≤ 125 ms), mid (125 ms < SDur ≤ 200 ms), and slow (SDur > 200 ms).

ceived as /p/ were rated higher than those that were categorized as /b/. Also, note that the misperception occurs at 10–30 ms VOTs, where the category boundaries are expected. This result supports the findings in experiment II that the listeners tend to perceive /b/ as /p/ at very fast rates.

In order to quantify these impressions and analyze the best exemplar for /b/ and /p/ for each listener and for each of the three rates, VOT values for highest rated /bi/ and /pi/ stimuli were tallied for each listener. When the highest rating was assigned to more than two stimuli with different VOT values, these VOT values were averaged. A two-way ANOVA was performed with consonant (/bi/ versus /pi/) and rate (fast versus mid versus slow) on the VOT of the highest rated stimulus for each listener. The results revealed a significant main effect of consonant $[F(1, 96) = 1057.6; p < 0.0001]$, such that best /p/'s have higher VOT than best /b/'s (mean VOT values for /b/ and /p/ were 14.2 and 38.8 ms, respectively). The main effect of rate was also significant $[F(2, 96) = 102.3; p < 0.0001]$, such that preferred VOT values were generally higher for slower rates (mean VOT values for fast, mid, and slow were 20.6, 25.2, and 33.7 ms, respectively). A significant interaction effect was also found $[F(2, 96) = 119.0; p < 0.0001]$, wherein the rate-dependent VOT changes on goodness judgments are larger for /p/ than /b/.

The goodness judgment results also showed a consistent asymmetrical relationship between /b/ and /p/. Mean goodness rating for correctly identified /p/ was higher for /p/ (7.59) than for /b/ (6.79), and the standard deviation (SD) was smaller for /p/ than for /b/ (SD: 2.80 ms for /b/ and 2.47 ms for /p/). On the other hand, the range of average VOT values for the highest rated tokens was narrower for /b/ than for /p/ (/b/: 9–20 ms, /p/: 19–61 ms). The reason for the perceptual asymmetry found in /b/ and /p/ is not clear, but it suggests that the listeners are more sensitive to the VOT differences for the best /b/ than for the best /p/. The result that the best /b/ tokens appeared in a narrower VOT range than the best /p/ tokens may be related with temporal psychophysics. If temporal processing of VOT differences is not linear, being particularly sensitive to VOT variation near zero, it might yield the result that the best /b/ tokens appeared in a narrower range of VOT than the best /p/ tokens. Alternatively, differences in phoneme frequency might also explain the perceptual asymmetry. The words with onset /b/ occur more frequently than the words with onset /p/ in the CVC context (Kessler and Treiman, 1997). More frequent access to the exemplars of word initial /b/ than the exemplars of word initial /p/ might make listeners to be more attentive to the detailed acoustic information for the best exemplars of /b/ than for /p/. More fastidious judgments toward /b/ might lower the overall goodness ratings for /b/. A third possibility is that onset /p/'s are simply more perceptually salient than onset /b/'s. Although the consonants /b/ and /p/ are rarely confused with each other, and the consonant /p/ is more confused with other consonants such as /t/ or /k/ in general (Miller and Nicely, 1955), it may not be the case for the highly ambiguous segments used in this study. Hence, subjects may simply find it easier to identify /p/'s, and so rate

them better. More work with different stop contrasts would be a likely way of pursuing these results further.

To summarize, then, what one finds in the goodness rating data is something like a hyperspace effect. The stimuli that were most often chosen as the best exemplars were not those that would most accurately characterize the distribution of the /p/ and /b/ categories, ones which fall in the center of the productions. Rather, the stimuli that are more extremely removed from other categories were rated more highly than their more moderate counterparts. Comparing the peak VOT values from Fig. 7 to the production distributions in Fig. 3 shows that listeners are selecting the top and bottom edges of the /p/ and /b/ distributions as best exemplars.

One other aspect of Fig. 7 is worth mentioning. The misidentified stimuli are consistently rated lower than the correctly identified ones, except the fast /b/ stimuli. That is, it seems that when subjects make an identification error, they are sensitive to information in the signal that runs counter to their misidentification. Particularly of note is that such misidentified tokens sometimes have VOT values that would lead one to miscategorize them (/b/ with long VOT, and /p/ with short VOT), but sometimes do not have such error-inducing VOT values. Hence, sometimes the VOT information could be cause for both miscategorization and the low rating, and sometimes only for the low rating.

The main results for production and perception can be summarized in Fig. 8. Figure 8 plots the category boundary based on the production results in experiment I (thin line) and the perceptual category boundary based on the perception results in experiment III (thick line). The average VOT values of the highest rated stimulus at fast, mid, and slow rates are also found in the same figure (filled circles for /b/ and unfilled triangles for /p/). The average VOT values of /b/ and /p/ stimuli at those three rates used in the analysis of experiment III were computed separately for /b/ stimuli (filled diamonds) and /p/ stimuli (unfilled squares) and superimposed on the figure. The rate effects can be clearly seen for the category boundaries both in production and perception as well as the best members of each category. Perception and production boundaries between voicing categories appeared in the same regions because of their close match. However, the perceptual boundaries exhibited a steeper slope because the perceptual boundaries shifted to the lower values at the fast rate because of the systematic and consistent misperception of fast rate /b/ tokens. The average VOT values of the highest rated stimuli appeared on both sides of the category boundaries in production. Similarly, the average VOT values of the best stimuli appeared on both sides of the perceptual boundaries, except the best /b/ at fast rate. This is again due to the misperception of /b/ at fast rates. The average VOT values of the best /b/ were shorter than the average VOT of the /b/ productions, just like expected. On the other hand, the average VOT values for the best rated /p/ were longer than the average VOT values for /p/ productions at mid and slow rates, but the average VOT for the best /p/ at fast rate was slightly shorter than the average VOT observed in the /p/ tokens at fast rates. The small VOT differences between average /p/ productions and the best /p/ were not expected. However, this was probably because the /p/ stimuli
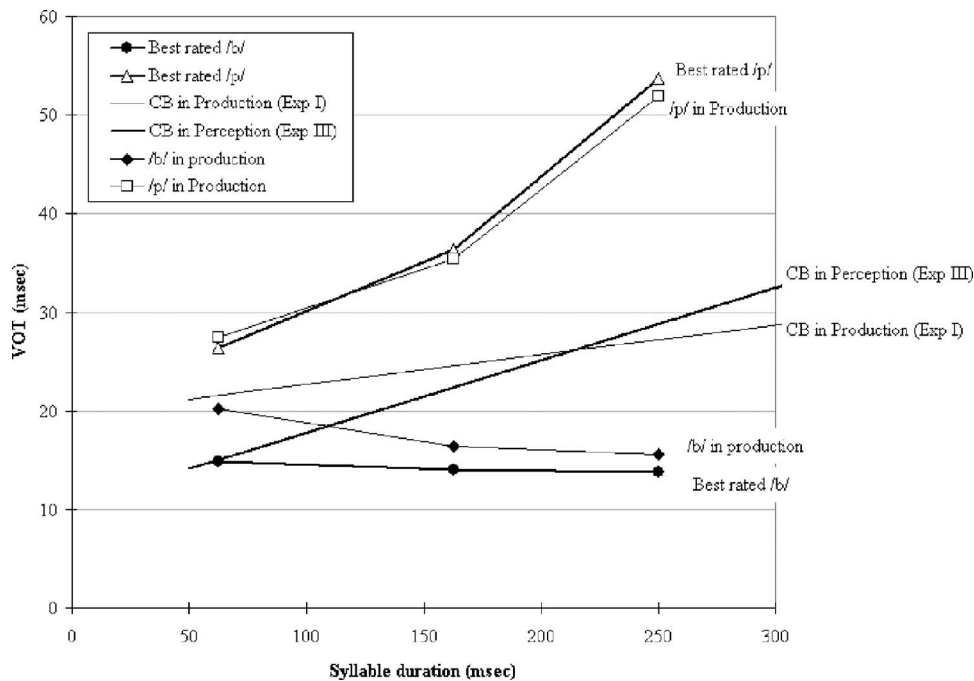
FIG. 8. The VOT values at the estimated category boundaries (CB) of /b/-/p/ productions in experiment 1, the VOT values at the perceptual CB estimated in experiment III, the average VOT values for the /b/ and /p/ tokens at three rates (fast, mid, and slow), and average VOT values for the best /b/ and /p/ stimuli at the three rates as a function of syllable duration. Three rates are defined by the average syllable durations (SDur) of the second and the third syllables in each stimulus: fast (SDur ≤ 125 ms), mid (125 ms < SDur ≤ 200 ms), and Slow (SDur > 200 ms).

in this study did not include alot of /p/ tokens with very long VOT. This might be a limitation of the use of natural stimuli. However, from the analysis of goodness ratings above, it was clear that the listeners selected the /p/ stimuli with longer VOT as the best /p/.

V. DISCUSSION

Using natural speech, the current data replicate the finding in studies on rate effects in voicing perception with synthesized and edited stimuli that the perceptual /b-p/ boundaries shift to longer VOT values as syllable durations increase (Miller and Volaitis, 1989; Volaitis and Miller, 1992). Thus, rate effects on perceptual identification found with synthesized and edited speech can be generalized to natural speech.

However, there are notable differences here from previous results concerning rate effects on voicing perception. While there were great differences between the location of the perceptual /b/-/p/ category boundary in Miller and Volaitis (1989) and Volaitis and Miller (1992) and the boundaries to differentiate /bi/ and /pi/ productions in Miller *et al.* (1986), the perceptual category boundaries estimated in the current perception experiments (experiments II and III) closely matched the production category boundaries estimated from the syllables in experiment I.

There are at least two likely reasons why the production-perception mismatch in VOT was smaller in the current investigation. Both are related to the nature of the stimuli and the degree to which voicing and rate information are encoded in the stimuli. First, unlike the previous studies, the current study employed naturally rate-varied stimuli, and therefore the current stimuli include more accurate information relevant to voicing perception than the ones used in previous studies. Hence listeners were better able to determine the intended productions of the speakers. Therefore, the boundary that distinguishes the two categories is very closely

aligned with actual produced differences except at fast speech rates. Another piece of evidence for the very close agreement of the speakers' and listeners' categories is found in the goodness judgment data. Even in cases where the listeners misperceived the intended voicing category of the speaker, the goodness evaluation data show that their misperception is only partial; such erroneously labeled tokens received systematically lower evaluations. The perceptual mechanism these results suggest will be further discussed below.

With respect to rate information, another notable difference in the current stimuli from those in previous studies was that each stimulus included three repetitive syllables instead of one isolated syllable. Even if the stimulus contains an ambiguous syllable, listeners could identify the consonant if one of the syllables in a stimulus provided the information that the listeners could use to identity the consonant clearly. The VOT and syllable durations of the three sequential syllables in each stimulus are never drastically different from one another. For example, VOT and syllable durations for the second and the third syllables in each stimulus differ only 0.5 ms for VOT and 6.8 ms for syllable duration. However, it is obvious that additional information was available to the listeners. The mismatch in VOT between production and perception boundaries might be reduced if the synthesized stimuli were repeated three times like the current study. It would be interesting to explore this possibility, but the authors expect that since VOT at the voicing category boundary tends to shift to longer VOT in long stimuli (e.g., Summerfield, 1981), the VOT at perception boundary could become longer as the stimuli become longer as well.

One note at this stage should be mentioned concerning the use of VOT as a criterial dimension for the English prevocalic voicing category. Although VOT is the dominant cue for voicing perception, other acoustic properties are known to influence listeners' voicing judgments [e.g., f0 shift

(Abramson and Lisker, 1985), formant transitions (Stevens and Klatt, 1974; Lisker, 1975; Summerfield and Haggard, 1977), consonant to vowel ratios (Port and Dalby, 1982), and amplitude of aspiration noise (Repp, 1979)]. Especially when they encounter ambiguous tokens, tokens appearing at around the category boundary on a VOT continua, listeners' identifications are demonstrably sensitive to some of these acoustic properties (Abramson and Lisker, 1985; Lisker, 1975; Whalen *et al.*, 1993). For some of the stimuli used in the current study, especially at fast rate stimuli, the produced /b/ and /p/ categories overlap in VOT. The fact that listeners correctly identify the consonant of these stimuli with wildly aberrant VOT values at above chance levels indicates that they are sensitive to some other properties of the stimuli besides VOT. This conclusion is also supported by the category goodness data. Tokens produced with VOT values in the wrong range, such as /b/ with VOT values greater than 20 ms, often were misidentified as /p/, however, receiving generally very low goodness ratings. These low ratings could simply mean that these stimuli were ambiguous with respect to the consonant, but also indicate some information conflicting with the listener's consonant identification as /p/ besides the VOT that indicates the intended /b/ category.

Although it seems reasonable that we would observe closer perceptual and production boundaries in cases where the stimuli exhibited all of the characteristics of naturally produced speech than in cases where the perceptual stimuli were artificial, this observation does not by itself indicate what sort of perceptual mechanisms might account for the listeners' performance, nor does it explain why previous studies should observe a specifically upward shift in the VOT boundary, relative to the current results. Of the four possible reasons given in the Introduction, the current results are compatible with the latter two, i.e., inclusion of an unnatural third category and the incompleteness of speech information. The current results suggest that some sort of hyperspace effect due to the use of artificial speech does not explain the production-perception mismatch in previous studies. The reason for this is that the current results for category goodness estimates tend to show the same sort of hyperspace effect as did Johnson *et al.* (1993). This hyperspace effect in the choice of best exemplars did not correspond to a shifted boundary in the identification task. Thus, we expect that the large mismatches in some of the previous studies were due either to task specific categories developed for the very wide range of VOT values for voiceless consonants or due to the synthetic and editing techniques being not entirely effective at conveying realistic differences in VOT.

The current study explored the notion of a hyperspace effect to explain the behaviors across subjects. In previous studies of the hyperspace effect (Johnson *et al.*, 1993; Whalen *et al.*, 2004a, b), the data analyses were conducted within individual subjects because the vowel space needs to be calibrated for each talker. This calibration problem would be reduced in the current analysis of English stops, because only the temporal dimension was examined. Newman (2003) examined the production-perception relationship within subjects and found a hyperspace effect for English stop consonant /p/ in a majority of subjects. Hence, as long as individu-

als exhibited a consistent and similar pattern in their behaviors, the hyperspace effect should appear in group data as a reflection of individual data. However, as we noted above, voicing contrasts are not solely determined by VOT. Analysis between the individual subjects' productions and perceptions might answer a question whether the hyperspace effect remains when the analysis is conducted in multiple dimensions (e.g., in both spectral and temporal domains).

Finally, concerning the perceptual mechanisms for rate that the current study suggests, we would contend that the current results do not fit well with a heavily abstractionist model, such as a traditional prototype model. To get rate effects with such prototypes, one must have an algorithmic extension of the evaluation procedure that matches stimuli to stored prototypes. However, one of the current results is that the rate effect is much larger with /p/ than with /b/. No simple scaling and matching algorithm will achieve this segment-specific effect, suggesting that a prototype model would have to either include a number of rate-specific prototypes or develop segment-specific scaling mechanisms. Note that switching from expressing VOT in absolute duration to a measure proportionate to the syllable duration will not solve this problem, since the rate variation in VOT is considerably less than proportional to the overall syllable duration (de Jong, 2001a), which would require reverse rate normalization. The good fit between production and perception obtained in the current study also highlights just how carefully these very specific algorithms would have to be tuned to the specifics of individual segments. In addition, the category goodness data suggest that such prototypes would have to be systematically shifted away from the distributional centers of the productions to get the hyperspace effect.

The very good match between production and perception suggests something more along the lines of "nontraditional" speech perception theories such as a direct realism view (Fowler, 1977, 1980) or an exemplar-based or episodic model of categorization (Nosofsky, 1986, 1992; Pisoni, 1997) where listeners retain fine phonetic details. Despite the high stimulus variability, listeners' accuracy of consonant identification did not decrease with rate except for the very fast stimuli, suggesting that listeners have knowledge of /b/-/p/ categories in their rate-varied details including VOT. The hyperspace effect also suggests the mechanisms employed in exemplar models (e.g., Nosofsky, 1986, 1992) where attentional weighting effectively expands dimensions which encode contrasts. Listeners are not just simply taking in the distributions, but apparently are sensitive to the orientation of contrast differences and what makes for an easily identifiable sound. Thus, such models allow for having both specific perceptual tuning to the detailed variation in the voicing categories and a specific mechanism that is sensitive to the abstraction of a linguistic contrast in voicing. A perceptual model that has a combination of tuning and abstraction seems to be what the current data call for.

However, an exemplar-based account with VOT-weighting does not fully explain the current perception results. At very fast rates (syllable duration < 100 ms), most of the /b/ stimuli were perceived as /p/. Since the talkers did not produce /b/ and /p/ in the same way even at fast rates (see de

Jong, 2001a), such systematic misperception of one category for the other under one condition would not be expected in our exemplar model, since there are plenty of exemplars in the region of space where the misperceptions are occurring. Yet, listeners persistently get the fast rate stimuli wrong, as though a decision boundary in the fast rate region was misplaced based on some independent motivation.

Such a misplacement might occur if the distributional patterns of the two categories are overgeneralized to the fastest rate conditions. One such possibility is that there is in fact some abstraction and normalization of rate effects. Since the distribution of rate effects of VOT values is more varied for /p/ than for /b/, the rate normalization effect for /p/ overestimates the shortening for /p/ at fast rates and induces a miscategorization of /b/. In favor of this sort of account, the slope of the categorization boundary from experiment III (in Fig. 8) is very nearly parallel to the shortening of VOT in the /p/ productions in experiment I. Hence, it is as if subjects were doing rate normalization without attending to the production of /b/'s. Since VOT does not vary as much by rate for /b/, the rate normalization effect would have to be somehow attached specifically to the /p/ category.

Though how to implement this model is not exactly clear, we should note that the general pattern of mismatch between production and perception is not likely to be limited to the voicing contrast in initial position. Our work with the rate varied data has also shown the same sort of systematic miscategorization for stops in postvocalic position as well, though in the opposite direction. Postvocalic voiceless stops in fast speech tend to be heard as voiced (de Jong *et al.* 2002). Similarly, our work with repetitive speech shows a systematic mismatch in the syllabic affiliation of stops whereby postvocalic stops tend to be categorized as prevocalic onsets (de Jong, 2001b, among others). While the main pattern of the current research indicates that the perceptual system is very closely tuned to productions, productions in the margins along dimensions such as speech rate exhibit pervasive mismatches with what the perceptual system is expecting.

VI. CONCLUSIONS

We examined rate normalization effects on voicing contrasts in prevocalic position using natural speech and found that rate normalization effects found with synthesized or modified speech also occur in the perception of natural speech. Speech rate affects both production and perception in a very similar manner. The perceptual identification system is neatly tuned to the distributions found in production. The results of goodness ratings also indicate that rate affects listeners' internal representations. Furthermore, the results of goodness ratings indicate that listeners store fine-grained information to distinguish voicing contrasts and are sensitive to the existence of contrasts along a particular dimension, such as VOT. Accurate identification of segments with aberrant VOT values suggests listeners use signal attributes in addition to VOT to differentiate the contrast. The general outcome of this perceptual tuning is that listeners' identifications of the voicing category are extremely good, even in the face of extreme variability. Listeners effectively deal with rate-induced variation in categorization tasks. However, in extremely fast rates, there are persistent mismatches between production and perception, which suggest that the perceptual system has some sort of generalization capacity that mischaracterizes the production variability in these margins. We suspect that examining such cases of production-perception mismatch provides an important insight into the nature of perceptual generalization in speech and has important and relatively poorly characterized implications for perceptual modeling as well as the function of speech communication systems. Of course, to do so requires us to examine speech perception more directly in light of actual speech production.

[1]When specific values are not available in the text, approximate values were read from figures. Mean values in Kessinger and Blumstein (1997) were estimated from Fig. 4, and their range values were estimated from Fig. 6. Syllable structures and the vowel following an initial consonant vary. Data on Kessinger and Blumstein (1998) was taken from the results of words "peak" and "peep." Allen and Miller's (1999) data were taken from the results of words "big" and "pig." For data from Lisker and Abramson (1964), values for prevoiced cases are not reported here. Newman (2003) used the syllable /pa/.

[2]Summerfield's (1981) data were taken from the Series B results in Experiment 6 and the results from the Series /bi-pi/ results in Experiment 6. Miller *et al.*'s (1984a) data were based on the results of Experiment 4a. Repp and Lin's (1989) data were based on the result of the isolated stimuli in Experiment III and the values reported in Repp and Lin (1991). In Repp and Lin (1989, 1991), initial periodic segments of natural production of /b/ were replaced by a periodic segment of equal duration of natural production of /p/. As for Repp and Lin (1991), data were taken from the result of the isolated stimuli. For Newman and Sawusch (1996), data were taken from the results of /blos/-/plos/ series in Experiment 5.

[3]Data for syllable durations longer than 300 ms were not plotted in Fig. 1 because there was only one sample.

Abramson, A. S., and Lisker, L. (**1985**). "Relative power of cues: F0 shift versus voice timing," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic Press, Orlando, FL), pp. 25–33.

Allen, J. S., and Miller, J. L. (**1999**). "Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words," J. Acoust. Soc. Am. **106**, 2031–2039.

Crystal, T. H., and House, A. S. (**1988**). "Segmental duration in connected speech signals: Current results," J. Acoust. Soc. Am. **83**, 1553–1573.

de Jong, K. (**2001a**). "Effects of syllable affiliation and consonant voicing on temporal adjustment in a repetitive speech-production task," J. Speech Lang. Hear. Res. **44**, 826–840.

de Jong, K. (**2001b**). "Rate-induced resyllabification revisited," Lang Speech **44**, 197–216.

de Jong, K., Nagao, K., and Lim, B. J. (**2002**). "The interaction of syllabification and voicing perception in American English," in *ZAS Papers in Linguistics*, edited by B. Pompino-Marschall and M. Rochon (Zentrum fur Allgemeine Sprachwissenschaft, Berlin), pp. 27–38.

de Jong, K. J., Lim, B. J., and Nagao, K. (**2004**). "The perception of syllable affiliation of singleton stops in repetitive speech," Lang Speech **47**, 241–266.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

K. Nagao and K. J. de Jong: Perceptual rate normalization 2897

Fowler, C. A. (**1977**). *Timing Control in Speech Production* (Indiana University Linguistics Club, Bloomington).

Fowler, C. A. (**1980**). "Coarticulation and theories of extrinsic timing," J. Phonetics **8**, 113–133.

Goldinger, S. D. (**1996**). "Words and voices: Episodic traces in spoken word identification and recognition memory," J. Exp. Psychol. Learn. Mem. Cogn. **22**, 1166–1183.

Goldman-Eisler, F. (**1968**). *Psycholinguistics: Experiments in Spontaneous Speech* (Academic, New York).

Johnson, K., Flemming, E., and Wright, R. (**1993**). "The hyperspace effect: Phonetic targets are hyperarticulated," Language **69**, 505–528.

Johnson, K., Flemming, E., and Wright, R. (**2004**). "Reply to Whalen *et al.*," Language **80**, 646–648.

Kessinger, R. H., and Blumstein, S. E. (**1997**). "Effects of speaking rate on voice-onset time in Thai, French, and English," J. Phonetics **25**, 143–168.

Kessinger, R. H., and Blumstein, S. E. (**1998**). "Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies," J. Phonetics **26**, 117–128.

Kessler, B., and Treiman, R. (**1997**). "Syllable structure and the distribution of phonemes in English syllables," J. Mem. Lang. **37**, 295–311.

Kidd, G. (**1989**). "Articulation-rate context effects in phoneme identification," J. Exp. Psychol. Hum. Percept. Perform. **15**, 736–748.

Lisker, L. (**1957**). "Closure duration and the intervocalic voiced-voiceless distinction in English," Language **33**, 42–49.

Lisker, L. (**1975**). "Is it VOT or a first-formant transition detector?" J. Acoust. Soc. Am. **57**, 1547–1551.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Lisker, L., and Abramson, A. S. (**1970**). "The voicing dimension: Some experiments in comparative phonetics," .in *Proceedings of the 6th International Congress of Phonetic Sciences*, edited by B. Hýýla, M. Romportl and P. Janota (Academia, Prague), pp. 563–567.

Magloire, J., and Green, K. P. (**1999**). "A cross-language comparison of speaking rate effects on the production of Voice Onset Time in English and Spanish," Phonetica **56**, 158–185.

Menard, S. (**2001**). *Applied Logistic Regression Analysis* (Sage, Thousand Oaks, CA).

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

Miller, J. L., and Baer, T. (**1983**). "Some effects of speaking rate on the production of /b/ and /w/," J. Acoust. Soc. Am. **73**, 1751–1755.

Miller, J. L., Dexter, E. R., and Pickard, K. A. (**1984a**). "Influence of speaking rate and lexical status on word identification," J. Acoust. Soc. Am. **76**, S89.

Miller, J. L., Green, K. P., and Reeves, A. (**1986**). "Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast," Phonetica **43**, 106–115.

Miller, J. L., Green, K. P., and Schermer, T. M. (**1984b**). "A distinction between the effects of sentential speaking rate and semantic congruity on word identification," Percept. Psychophys. **36**, 329–337.

Miller, J. L., Grosjean, F., and Lomanto, C. (**1984c**). "Articulation rate and its variability in spontaneous speech: A reanalysis and some implications," Phonetica **41**, 215–225.

Miller, J. L., and Volaitis, L. E. (**1989**). "Effect of speaking rate on the perceptual structure of a phonetic category," Percept. Psychophys. **46**, 505–512.

Newman, R. S. (**2003**). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," J. Acoust. Soc. Am. **113**, 2850–2860.

Newman, R. S., and Sawusch, J. R. (**1996**). "Perceptual normalization for speaking rate: Effects of temporal distance," Percept. Psychophys. **58**, 540–560.

Nosofsky, R. M. (**1986**). "Attention, similarity, and the identification-categorization relationship," J. Exp. Psychol. Gen. **115**, 39–57.

Nosofsky, R. M. (**1992**). "Exemplar-based approach to relating categorization, identification, and recognition," in *Multidimensional Models of Perception and Cognition*, edited by F. G. Ashby (Erlbaum, Hillsdale, NJ), pp. 363–393.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllabic nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Pisoni, D. B. (**1997**). "Some Thoughts on 'Normalization' in Speech Perception," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego), pp. 9–32.

Pisoni, D. B., and Tash, J. (**1974**). "Reaction time to comparisons within and across phonetic categories," Percept. Psychophys. **15**, 285–290

Port, R. F., and Dalby, J. (**1982**). "Consonant/vowel ratio as a cue for voicing in English," Percept. Psychophys. **32**, 141–152.

Repp, B. H. (**1979**). "Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants," Lang Speech **22**, 173–189.

Repp, B. H., and Lin, H. B. (**1989**). "Effects of preceding context on discrimination of voice onset times," Percept. Psychophys. **45**, 323–332.

Repp, B. H., and Lin, H. B. (**1991**). "Effects of preceding context on the voice-onset-time category boundary," J. Exp. Psychol. Hum. Percept. Perform. **17**, 289–302.

Roach, P. (**1998**). "Some languages are spoken more quickly than others," in *Language Myths*, edited by L. Bauer and P. Trudgill (Penguin, London), pp. 150–158.

Sawusch, J. R., and Newman, R. S. (**2000**). "Perceptual normalization for speaking rate II: Effects of signal discontinuities," Percept. Psychophys. **62**, 285–300.

Stevens, K. N., and Klatt, D. H. (**1974**). "Role of formant transitions in the voiced-voiceless distinction for stops," J. Acoust. Soc. Am. **55**, 653–659.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Summerfield, Q. (**1981**). "Articulatory rate and perceptual constancy in phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.

Summerfield, Q., and Haggard, M. (**1977**). "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants," J. Acoust. Soc. Am. **62**, 435–448.

Utman, J. A. (**1998**). "Effects of local speaking rate context on the perception of voice-onset time in initial stop consonants," J. Acoust. Soc. Am. **103**, 1640–1653.

Volaitis, L. E., and Miller, J. L. (**1992**). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," J. Acoust. Soc. Am. **92**, 723–735.

Wayland, S. C., Miller, J. L., and Volaitis, L. E. (**1994**). "The influence of sentential speaking rate on the internal structure of phonetic categories," J. Acoust. Soc. Am. **95**, 2694–2701.

Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "$F0$ gives information even with ambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.

Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., and Iskarous, K. (**2004a**). "Vowel production and perception: Hyperarticulation without a hyperspace effect," Lang Speech **47**, 155–174.

Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., and Iskarous, K. (**2004b**). "Vowel targets without a hyperspace effect," Language **80**, 377–378.

Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., and Iskarous, K. (**2004c**). "Whalen *et al.* reply," Language **80**, 648–649.

Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters

Pierre A. Hallé[a]

Laboratoire de Phonétique et Phonologie, 19 rue des Bernardins, 75005 Paris, France,
Laboratoire de Psychologie et Neurosciences Cognitives, 71 Av. Edouard Vaillant,
92774 Boulogne-Billancourt, France and Haskins Laboratories, 300 George Street, New Haven, CT 06511

Catherine T. Best[b]

MARCS Auditory Laboratories, University of Western Sydney, Penrith South DC, NSW 1797, Australia
and Haskins Laboratories, 300 George Street, New Haven, CT 06511

French listeners perceive illegal /tl/ and /dl/ clusters as legal /kl/ and /gl/, suggesting that /dl, tl/ undergo "phonotactic perceptual assimilation" to the phonetically most similar permissible clusters [Hallé *et al.*, J. Exp. Psychol. Hum. Percept. Perform. (1998)]. However, without a comparison to native speakers of a language allowing initial /tl, dl/, other explanations remain open (e.g., universal phonetic biases). Experiment 1 compared native French and Hebrew listeners on perception of Hebrew /tl/-/kl/ and /dl/-/gl/. On a language-specific phonotactics account, these contrasts should be difficult for listeners whose language disallows initial /tl, dl/ while allowing /kl, gl/ (French), but not for listeners whose language permits all four clusters (Hebrew). Indeed, French but not Hebrew listeners showed difficulty discriminating /tl/-/kl/, and tended to categorize the initial consonant of /tl/ as /k/; analogous effects for /dl/-/gl/ were weaker. Experiment 2 tested speakers of American English, which also disallows initial /tl, dl/ but realizes stop-voicing differently than French or Hebrew, to examine possible contributions of language-specific phonetic settings. Their performance was similar to that of French listeners, though they had significantly greater difficulty with /dl-/gl/. The results support the proposal of language-specific phonotactic perceptual assimilation, with modest contributions from language-specific phonetic settings.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2534656]

I. INTRODUCTION

In the early 1930s, Polivanov (1931) suggested that people's native language phonological system must bias the way they perceive foreign sounds. Thereafter, metaphors such as "phonological deafness" and "phonological filter" coined in the 1930's (cf. Trubetzkoy, 1939) have been widely used and tacitly accepted as correct characterizations of how listeners perceive non-native sounds. They reflect Polivanov's (1931) proposition that the native phonological system molds perception of non-native speech patterns to follow its rules.

In that proposition, Polivanov included the phoneme "grouping laws" [Polivanov (1931), p. 80] that govern permissible sound sequences. For instance, he noted the case of Japanese listeners perceiving the word "drama" as *dorama*, or alternatively as *zurama*, illustrating how the specific phonotactics of a language—here, the general ban against consonant clusters in Japanese—may constrain the way non-native sequences are perceived. This example can be viewed as an early illustration of what was later called "phonological repair" in both loanword phonology and cross-language speech perception research.[1] The subsequent findings in both lines of research are largely consistent with Polivanov's general intuition and informal observations. Indeed, it seems that the knowledge of which speech sounds and sound combinations occur in real-life utterances does guide and facilitate speech perception, at the predictable expense of precision in the phonetic analysis of ill-formed utterances.

There still are a number of unresolved questions about *perceptual* phonological repairs (as opposed to repairs in *produced* loanword adaptations). The first question concerns the relative roles of low (phonetic/acoustic) and higher (phonological) levels of speech analysis. This question is the subject of an ongoing debate in loanword phonology between proponents of exclusively phonological motivations for loanword adaptations (Paradis and LaCharité, 1997, 2001) and proponents of the "phonetic approximation" view (see, among others, Shinohara, 2006; Silverman, 1992; Vendelin and Peperkamp, 2004; Yip, 1993). In the broader field of speech perception, the phonological versus phonetic tension is directly related to the issue of language-specificity versus universality, respectively. Are language-specific phonological rules and constraints so firmly imprinted in listeners' knowledge of the native system that they supersede physical evidence? Are certain physical characteristics of speech sounds so salient that potential repairs are universally ruled out?

Another issue is that of the repair of ill-formed structures above the segmental level, for sequences of segments. It has often been discussed in the loanword phonology litera-

[a] Electronic mail: pierre.halle@univ-paris5.fr
[b] Electronic mail: c.best@uws.edu.au

ture (e.g., Paradis and LaCharité, 1997) but has seldom been addressed in the speech perception literature, with the notable exception of the well-documented case of the epenthetic "illusory" vowel heard by Japanese listeners presented with consonant clusters. Dupoux and colleagues examined the perception of utterance-medial consonant clusters by Japanese versus French listeners (Dehaene-Lambertz *et al.*, 2000; Dupoux *et al.*, 1999, 2001). They found that Japanese, but not French listeners, hear the same epenthetic /u/ in nonwords such as *ebza*, perceived as /ebuza/, and in pseudowords such as *sokdo* or *mikdo* (from *sokudo* "speed" or *mikado* "emperor"). This is precisely the type of phonological repair Polivanov had observed Japanese listeners to apply to words such as *drama*.[2] In several studies, Dupoux and colleagues used systematic cross-language comparisons to reveal the language-specific nature of the repair. Their findings show that language-specific phonological constraints can supersede physical evidence. Presence/absence of a vowel is prima facie a large phonetic difference, yet because of their phonological system, Japanese listeners perceive a vowel that is physically absent. This failure to perceive the large phonetic difference between presence/absence of a vowel is all the more remarkable in that vowel epenthesis introduces a change in syllabic structure, from the perspective of languages which allow word-medial consonant clusters.

Hallé *et al.* (1998) reported another case of perceptual phonological repair associated with native phonotactic constraints but involving consonant substitution rather than vowel insertion. In this case, the phonotactic repair leaves syllabic structure unchanged. The reported data showed that French listeners strongly tend to perceive utterance-initial /tl/ and /dl/, which are not permissible in French, respectively, as /kl/ and /gl/, which are permissible. This "dental-to-velar shift" effect (in the formulation of Hallé *et al.*) was found to be stronger for /tl/ than /dl/. It was demonstrated using mainly identification tasks on nonwords such as *dlopta* or *tlabod*, pronounced by a French speaker, as compared to nonwords such as *dropta* or *trabod*, whose word-initial /dr, tr/ clusters are permissible in French. The latter nonwords did not induce any dental-to-velar shift. However, this study used a within-language rather than a cross-language comparison, leaving unresolved the question of whether the repair is language-specific rather than universal. Dupoux *et al.* (2001) commented that the dental-to-velar shift in Hallé *et al.* (1998) could be "due to universal effects of compensation for coarticulation" and that "it could be that /dl/ is universally harder to perceive than /gl/." In a nutshell, the compensation for coarticulation argument holds that listeners expect dorsal stops to be fronted in the context of the coronal contact for /l/ (Mann, 1980; Mann and Repp, 1981). Thus, a stop with an articulatory and hence acoustic fronted quality can be heard as a fronted dorsal stop in /l/ context, resulting in a dental-to-velar perceptual shift. To evaluate whether the /dl, tl/→/gl, kl/ perceptual shift reflects such universal perceptual processing or, rather, a language-specific phonotactic repair, a cross-language comparison is needed. If the effect is language-specific, the dental-to-velar phonotactic perceptual repair should occur only in listeners whose language disal-

lows dental stop +/l/ clusters, and not in listeners whose language permits these clusters.

It is the need for a cross-language examination of the dental-to-velar phonotactic repair that is the primary motivation for the present study. A comparison language that permits /dl, tl/ is needed to determine whether the repair is specific to the phonotactic constraints of the listener's language, or rather is associated with universal perceptual processes (e.g., compensation for coarticulation). French listeners' tendency to identify /dl/ as /gl/ and /tl/ as /kl/ clearly reflects perceptual confusability between /tl/ and /kl/ and between /dl/ and /gl/; this should entail poor discrimination of /dl/-/gl/ and /tl/-kl/ contrasts. We therefore evaluated discrimination of these clusters by French listeners. For the critical cross-language examinations, the comparison language had to allow /gl, kl/ as well as /dl, tl/ onset clusters. If language-specific phonotactic repair is responsible for the French perceptual bias, then, the native listeners of the comparison language should have little or no difficulty discriminating /dl/-/gl/ and /tl/-/kl/. But if universal perceptual constraints are instead responsible, a possibility suggested in Dupoux *et al.* (2001), then even the listeners of the comparison language should display substantial difficulty perceiving these contrasts.

Only a few languages have coronal-dorsal contrasts such as /tl/-/kl/, which, in itself, might reflect a universal trend toward avoiding these contrasts. One language that does have these contrasts is Modern Hebrew. It allows all possible stop+liquid clusters word-initially (Rosen, 1962; Téné, 1972), including /dl, tl/ and /gl, kl/, as well as /dr, tr/ and /gr, kr/. The voiced stops of Hebrew have substantial voicing lead (Laufer, 1998; Raphael *et al.*, 1995). The voiceless stops have been reported as intermediate between phonetically short lag unaspirated and long lag aspirated (Obler, 1982; Raphael *et al.* 1983, 1995). The phonological voicing distinction for stops is thus similar though not identical in Hebrew and in French, which contrasts prevoiced and short lag unaspirated stops (see, for example, Nearey and Rochet, 1994). French and Hebrew have phonetically similar "light" /l/s (French: Chafcouloff, 1979; Simon, 1967; Hebrew: Chayen, 1973). Their /r/s have also been described as similar in place (uvular), although they might differ somewhat in manner of articulation: approximant or fricative /r/s in the French of our Parisian listener group (cf. Hallé *et al.*, 1999) versus approximant or trilled /r/s in Hebrew[3] (Devens, 1978, 1980; Laufer, 1990; Rosen, 1962). Therefore, in Experiment 1, we compared native French and Hebrew listeners on their discrimination of the Hebrew dental-velar contrast in the context of a following /l/ or /r/; Experiment 2 extended these cross-language comparisons to American English, which also disallows /dl/ and /tl/ but differs from French in the phonetic realizations of its coronal stops and its liquids. Would these phonetic differences have substantive impact on American relative to French listeners' perception of /dl, tl/? Given that all stop+/r/ clusters are permissible in the three languages (French, English, Hebrew), the /r/ context was used for comparison with the study of Hallé *et al.* (1998), in which stop +/r/ clusters served as a within language baseline comparison.

P. A. Hallé and C. T. Best: Dental-to-velar perceptual assimilation

While the cross-language comparison is the main motivation for this study, there are additional reasons to reconsider the earlier findings. First, there may be reason for concern about the phonetic and/or articulatory quality of the illegal /dl, tl/ speech tokens used in the earlier study. They were produced by a native speaker of French, who had no familiarity with any language allowing /dl, tl/ word-initially and had no training in phonetics. The critical stimuli thus could simply have been mispronounced as velar (or ambiguous) stop+/l/. Hallé *et al.* (1998) addressed this concern in two ways. They compared the acoustic properties of the stop release bursts of the /dl, tl/ versus /gl, kl/ clusters. Spectral cues to place of articulation were found to be more dental-like in the intended dentals than in the intended velars, in agreement with classic measures of the spectral properties of stop bursts (Halle *et al.*, 1957; Kewley-Port, 1983; Stevens and Blumstein, 1978). These properties should provide sufficient information for reliable perception of place of articulation from bursts alone (Blumstein and Stevens, 1980). Hallé *et al.* (1998) then conducted a "phonetic gating" experiment and found that French listeners heard a dental rather than a velar stop in the shortest fragments, which corresponded to the stop bursts alone. It was thus argued that the initial stops of /dl, tl/ sufficiently met the requirements for dental place of articulation. It is possible, however, that the French speaker of Hallé *et al.* (1998) produced ambiguous stops, intermediate between velar and dental, in the /dl, tl/ clusters. Indeed, assuming that speech perception and production are strongly linked (cf. Fowler *et al.*, 2003, for an indepth discussion), the dental-to-velar shift could conceivably occur in the French speakers' *productions* of /dl, tl/. These potential stimulus shortcomings are readily avoided by recording a native speaker of Hebrew, who must be able to produce differentiated velar- and dental-stop +/l/ clusters in word-initial position, because they are needed for contrasting minimal-pair words such as *tlulim-klulim* ("steep"–"included").

A second important concern is that of possible lexical influences. Although the nonwords in the study of Hallé *et al.* (1998), such as *dlopta* or *tlabdo*, had been designed to bear little similarity with existing words, use of nonwords does not prevent lexical access processes from operating. The /dl, tl/ nonwords could conceivably activate words with phonetically similar onsets, such as /gl, kl/. Such lexical neighborhood activation was plausible given that the nonwords used were quite word-like in terms of syllabic pattern and phonotactic probabilities (except of course for the /dl, tl/ onsets), and were mixed with a wealth of very French-like nonword fillers. The perceptual repair of /dl, tl/ does not appear to be *strictly* lexically driven because, although French words beginning with /p1, bl/+/a, o, ɔ/ (the vowel context used in Hallé *et al.*, 1998) are as frequent as words beginning with /gl, kl/+/a, o, ɔ/, according to the "LEXIQUE" database (New *et al.*, 2001) (/pl, bl/ words vs /kl, gl/ words: 438 vs 536 types; 1124 vs 962 per million cumulative frequencies), there were exceedingly few dental-to-labial repairs. Thus, a strict lexical account of the /tl, dl/→/kl, gl/ perceptual repair does not work. However, the dental-to-velar shift could nevertheless be lexically *biased* to some

extent. Specifically, the stronger dental-to-velar shift found for /tl/ than for /dl/ is compatible with the higher frequency of /k/- than /g/-initial words in the /l/+/a, o, ɔ/ context (/kl/ vs /gl/ words: 369 vs 167 types, 715 vs 247 per million cumulated frequency). The cross-linguistic approach we employ here, using classic discrimination and categorization tasks with Hebrew stimuli, should help solve this puzzle. The Hebrew stimuli sound foreign rather than French-like, which should encourage French listeners to attend to their phonetic rather than lexical properties, thus minimizing the likelihood of lexical influences. To further minimize lexical effects, monosyllabic stimuli were used, because they are presumably less likely to suggest lexical items to French listeners than the multisyllabic stimuli of the earlier study (Hallé *et al.*, 1998).

II. EXPERIMENT 1

We first examined the perception of Hebrew /dl, tl/ clusters by native listeners of French versus native speakers of Israeli Hebrew. If the dental-to-velar perceptual shift holds for French listeners presented with Hebrew stimuli, they should perceive Hebrew /dl, tl/ and /gl, kl/ as similar and thus have trouble discriminating the Hebrew /dl/-/gl/ and /tl/-/kl/ contrasts. The Hebrew /dr/-/gr/ and /tr/-/kr/ contrasts were used as control baseline contrasts to gauge French listeners' performance at perceiving Hebrew dental-velar contrasts in permissible, contrastive stop+liquid clusters, which French listeners should label and discriminate quite well. Hebrew listeners should encounter little difficulty with all the dental-velar stop+liquid cluster contrasts, whether with /r/ or with /l/, because they are all legal and contrastive in Hebrew. Yet, were the dental-to-velar perceptual shift for /dl, tl/ a universal tendency, Hebrew listeners could show difficulty discriminating /dl/-/gl/ and/or /tl/-/kl/.

A. Method

1. Participants

Twelve native French-speaking students at Paris V University (mean age 22, age range 19–25) and eleven native speakers of Hebrew recruited in Paris (mean age 23, age range 21–29), participated in the experiment for a small participation payment. The data for one additional French participant were not retained because of failure to participate in the labeling part of the experiment. The data for one additional Hebrew participant were not retained because she had been residing in France for more than 20 years (the other participants were in Paris on a short stay) and was also well outside the age range of the other participants.

2. Stimuli

Twenty-four CCV monosyllabic items were constructed by crossing the vowels /a, i, u/ with the clusters /dl, tl, gl, kl, dr, tr, gr, kr/, which are all legal word-initially in Hebrew. The clusters were thus composed of a dental or a velar stop, voiced or voiceless, followed by /l/ or /r/. The three cardinal vowels /a, i, u/ were chosen so that the full extent of the Hebrew vowel space (/a, e, i, u, o/) would be represented. A randomized list, containing eight tokens of each item, was

TABLE I. Characteristics of $F0$ contours in the speech materials retained: $F0$ at contour onset, $F0$ range, mean $F0$ (all in Hz), and mean $|d^2F0/d^2t|$ (an index of $F0$ fluctuations),[a] and contour duration (in ms), according to cluster.

Cluster	Onset $F0$	$F0$ range	Mean $F0$	$|d^2F0/d^2t|$	Duration
/tl/	147	15	152	5.6	258
/kl/	150	14	151	6.7	249
/tr/	146	19	151	8.7	280
/kr/	147	22	151	10.9	272
/dl/	119	43	139	17.4	385
/gl/	121	42	138	13.6	381
/dr/	124	37	138	28.2	397
/gr/	124	42	138	21.5	386

[a]See Hallé et al. (1991, pp. 303–304) for further details on this fluctuation index.

read aloud by a male native speaker of Hebrew who resides in Israel. The recording was made in an anechoic chamber, using a professional quality microphone and a DAT tape recorder. The speech materials were then digitized (16 kHz sampling rate, 16 bit resolution) and transferred to individual audio files. Two native speakers of Hebrew, students in phonetics, judged whether each stimulus had been correctly pronounced by checking them against a list of the intended pronunciations transcribed in Hebrew script. All the tokens were approved by both judges and therefore retained for selection as experimental stimuli.

For each of the 24 stimulus syllable types (4 stops × 2 liquids × 3 vowels), four tokens (out of eight repetitions) were selected for use in the perceptual experiments, for which the prosodic characteristics (syllable duration, $F0$ contour, and loudness) were balanced as well as possible within each set of items to be compared (e.g., the /gla/ vs /dla/ set). In order to enhance prosodic homogeneity, peak intensities in

the vowel portion were equalized within and across these sets. The prosodic features of the selected materials are shown in Table I.

The phonetic-acoustic characteristics of the clusters used, in particular of the initial stops, were critical to know for the velar-dental distinctions tested. Because few phonetic-acoustic data are available for Hebrew consonants and clusters, we acoustically analyzed the stimuli to be used in the planned perceptual tasks. In the stimuli produced by our speaker, the voiced stops had substantial prerelease voicing, whereas the voiceless stops had substantial postrelease aspiration, thereby somewhat departing from the medium lag VOT that has been reported for Hebrew (e.g., Raphael et al., 1995). Thus, the voiced stops were prevoiced while the voiceless stops were long-lag aspirated voiceless.

Three main characteristics distinguished velars and dentals. They are detailed in Table II. First, the spectral center of gravity (henceforth, SCG) in the release burst portion was lower overall for velars than for dentals by about 250 Hz, $F(1,72)=41.16$, $p<0.00001$, but the difference was more marked for the /r/ than the /l/ context (333 vs 176 Hz), as suggested by a significant interaction between liquid context (/l/ vs /r/) and place of articulation (dental versus velar), $F(1,72)=3.91$, $p<0.05$. There was some variability, especially with regard to vowel context and voicing, as can be seen in Table II: SCG for /dl, tl/ was much lower in the -/u/ than in the -/i/ or -/a/ contexts (3.2 vs 4 or 4.2 kHz) while SCG for /gl, kl/ was much less affected by the -/u/ context and remained in the 3.4–3.8 kHz range. As a result, SCG values for /dlu, tlu/ (3.2 kHz) are more typical of velar than dental place, whereas those for /dli, tli/ or /dla, tla/ (~4 kHz) are more typical of dental than velar place. This would predict more dental-velar shifts for /dlu, tlu/ than for /dli, tli/ or /dla, tla/, if SCG is an important cue in perception. For all other contexts, SCG was always lower for velars than dentals. Within the /l/ context, the difference was numerically, although not statistically larger for /dl/-/gl/ (250 Hz)

TABLE II. Acoustic characteristics of the dental- vs velar-initial cluster stimuli according to liquid and vowel context. Voice onset time (VOT) is expressed in ms, spectral center of gravity (SCG) in Hz, and burst integrated energy (BIE) in dB s. Δs stand for (velar - dental) differences.

| | | /l/ context | | | | | | /r/ context | | | | | |
| | | Voiced | | | Voiceless | | | Voiced | | | Voiceless | | |
	Vowel	/dl/	/gl/	Δ	/tl/	/kl/	Δ	/dr/	/gr/	Δ	/tr/	/kr/	Δ
VOT	/a/	−147	−97	*50*	49	78	*29*	−97	−85	*12*	40	99	*59*
	/i/	−126	−105	*21*	60	101	*41*	−94	−76	*18*	51	101	*50*
	/u/	−106	−94	*12*	67	93	*26*	−102	−97	*5*	60	95	*35*
	mean	−126	−99	*27*	59	91	*32*	−98	−86	*12*	50	98	*48*
SCG	/a/	4212	3552	*−660*	4066	3699	*−367*	3944	3631	*−313*	3958	3724	*−234*
	/i/	4017	3719	*−298*	3963	3858	*−105*	3939	3340	*−599*	3915	3566	*−349*
	/u/	3206	3413	*+207*	3571	3742	*+171*	3833	3699	*−134*	3879	3518	−361
	mean	3812	3561	*−251*	3867	3767	*−100*	3906	3557	*−349*	3918	3602	*−316*
BIE	/a/	1.25	1.49	*0.24*	3.17	4.72	*1.55*	1.13	1.81	*0.69*	2.57	5.57	*3.00*
	/i/	1.30	1.55	*0.25*	3.86	6.22	*2.36*	1.22	1.93	*0.72*	3.47	5.82	*2.36*
	u	1.32	1.80	*0.47*	4.22	5.91	*1.70*	1.35	2.15	*0.80*	3.67	5.90	*2.22*
	mean	1.29	1.61	*0.32*	3.75	5.62	*1.87*	1.23	1.97	*0.73*	3.24	5.76	*2.53*

P. A. Hallé and C. T. Best: Dental-to-velar perceptual assimilation

than /tl/-/kl/ (100 Hz), $t(22) < 1$. Second, integrated energy in the burst was greater for velars than for dentals (3.74 vs 2.38 dB s), $F(1,72)=172.75$, $p < 0.00001$. This measure is interesting because perception of loudness depends on energy integrated over time, at least for rather short fragments of speech signal (Hughes, 1946; Scharf, 1978). It produced much smaller values for voiced than voiceless stops (1.53 vs 4.59 dB s), $F(1,72)=875.50$, $p < 0.00001$, in keeping with the observation that voiced stops have quieter bursts than voiceless stops (Zue, 1976). Not surprisingly, then, the velar-dental differentials for this measure are smaller for voiced than voiceless stops (0.53 vs 2.20 dB s), as suggested by the Place $\times$ Voice interaction, $F(1,72)=64.93$, $p < 0.00001$. Third, VOTs were significantly longer for velars than for dentals, consistent with the literature on VOT variation according to place of articulation (Fischer-Jørgensen, 1954; Lisker and Abramson, 1964; Nearey and Rochet, 1994; Peterson and Lehiste, 1960; Saerens *et al.*, 1989; see Cho and Ladefoged, 1999, for an overview). For voiceless stops, VOTs were longer for velars than dentals (95 vs 55 ms on average) in both /l/ and /r/ contexts, $F(1,36)=205.5$, $p < 0.00001$. Voiced stops were all produced with voicing lead (i.e., negative VOTs). Voicing leads were significantly shorter for velars than dentals (93 vs 112 ms on average), $F(1,36)=12.43$, $p < 0.005$.

Dental clusters did not differ from velar clusters with respect to other measured acoustic cues. The duration of the liquid component was 115 vs 113 ms for /tl/ vs /kl/, 122 vs 135 ms for /dl/ vs /gl/, 114 ms for both /tr/ and /kr/, and 134 vs 124 ms for /dr/ vs /gr/ (all nonsignificant pairwise differences). The /r/s of the C+/r/ clusters were realized as uvular approximants or, more often, as uvular trills (cf. Devens, 1980), which occurred half of the time after voiced stops and virtually always after voiceless stops, regardless of place of articulation of the initial stop. The first three formants of the vowel following the liquid (/l/ or /r/) were measured at 25%, 50%, and 75% of the vowel duration. The formant patterns were virtually identical in all dental-velar pairs (less than 50 Hz differences). They were conditioned only by the vowel itself and by the preceding liquid, not by the initial stop. $F2$ had a lower locus after /r/ than after /l/ ($\sim$1 vs 1.9 kHz), consistent with the literature on French /l/ vs /r/ (Chafcouloff, 1979). Analogously, the three first formants measured midway through the steady state of the liquids /l/ and /r/ only depended on the liquid itself and the following vowel, not the place of articulation of the initial stop.

Lexical biases in French (Experiment 1) or American English (Experiment 2) listeners' perception are prima facie unlikely with stimuli that were produced by a native Hebrew speaker, and indeed sounded foreign to both French and American listeners. However, among the six /kl, gl/ syllables, three could induce a lexical bias in French: /klu/, /glu/, and /gla/ resemble a French word [*clou* "nail," *glou* onomatopoeic, and *glas* "(bell) toll"]; /klu/, /glu/, and /gli/ resemble an English word (*clue*, *glue*, and *glee*). These individual words could attract more dental-to-velar shifts in the matched /dl, tl/ syllables than in the others, and result, for example, in lower discrimination performance for /tlu/-/klu/, /dlu/-/glu/, and /dla/-/gla/ than other dental-velar contrasts by

French listeners. Aside from individual words, lexical bias might be mediated by the cohorts of words beginning with /kl, gl/. Among these, in both French and English, words with /kla/ or /gla/ as initial syllable are by far the most frequent; those with /klu/ or /glu/ are the least frequent; those with /kli/ or /gli/ fall in between.[4] Thus, a cohort-mediated lexical bias would yield the poorest discrimination performance for the /a/ vowel context and the best one for /u/ in both French and English.

3. AXB discrimination task

The selected stimulus tokens were used to make up categorial *AXB* discrimination triads, in which *A* and *B* differed with respect to place (velar versus dental) and *X* served as the target item, to be judged as to its category match with *A* or *B*. It was always a categorial match, in that it was always a different token than the *A* or *B* item that it matched phonologically. For each of the four possible triad orders (*AAB*, *ABB*, *BAA*, and *BBA*), eight triplets were constructed (hence 32 combinations) in such a way that each token appeared equiprobably in each position and was not repeated within a given triplet. This yielded 384 triplets (32 combinations $\times$3 vowels $\times$2 voicings $\times$2 liquids), hence 384 test trials. The trials were presented in random order, blocked by 12 trials. The interstimulus interval was set to 1 s, the intertrial interval to 4 s, and the interblock interval to 8 s. Participants were allowed to pause midway during the test phase. The test phase was preceded by a training phase of ten *AXB* trials, which did not appear in the test phase. Three trials were "easy" in that *A* and *B* differed in both initial consonant *and* liquid. No feedback was given to participants in either the training or the test phase.

For each trial, participants of both language groups were asked to indicate whether they perceived *X* as similar to *A* or to *B* by pressing one of two buttons labeled "1" and "3" (for *A* and *B*). They were instructed to respond on each trial, even if they had to guess, and to respond as fast as possible, as soon as they were confident of their response. There was no constraint that they wait until the third stimulus *B*. Hence, when listeners are quite confident, they may respond even before the final item of the triad, actually performing an *AX*-like speeded task in this case. We call this variant of the *AXB* discrimination procedure the "free RT" speeded *AXB* paradigm. The response times (RTs) were measured from the release burst of the initial consonant of the third stimulus. Thus, early responses (i.e., occurring before the third item of the triad) have negative RTs. This procedure has proved useful to enhance RT differences between listener groups or between stimulus contrasts (Hallé *et al.*, 2004). One potential drawback is that within-group RT variability may be large.

4. Categorization task

Only French participants had to categorize the Hebrew clusters in terms of French segmental categories. The task was not administered to Hebrew participants, given that all stimuli had been found to match intended pronunciation by two phoneticians, native speakers of Hebrew. In the test phase, each of the six /dl/ and /tl/ syllable types (2 clusters $\times$
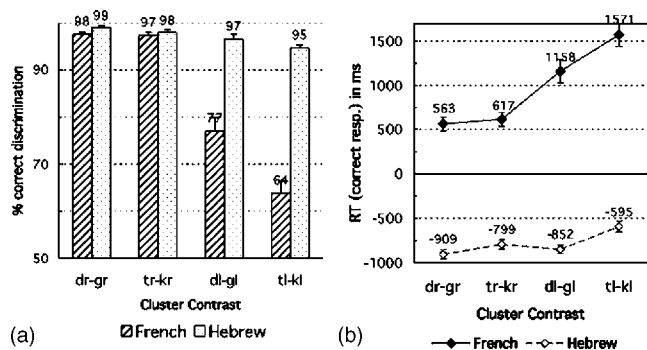
FIG. 1. Discrimination performance of Hebrew vs French subjects: (a) Percent correct discrimination, and (b) RT for correct responses in ms. Error bars indicate standard error.

FIG. 2. Categorization of the initial stop of Hebrew /dl/ and /tl/ by French subjects: (a) Percentages of dental and velar responses, and (b) corresponding "fit indexes." Error bars indicate standard error.

3 vowels) was used eight times (4 tokens × 2 repetitions), making 48 trials. The other 18 syllable types (6 clusters × 3 vowels) were used only four times (4 tokens), because they were presumably easier, making 72 trials. The test phase thus consisted of 120 trials. It was preceded by a training phase of 16 trials, which included two /dl/ and two /tl/ trials. No feedback was given to participants in either the training or the test phase.

For each trial, participants were first presented with a given syllable twice in a row. They were instructed to categorize its *initial consonant* by choosing one of ten consonants illustrated with the help of rhyming French keywords displayed on the screen (*paon*, *temps*, *Caen*, *banc*, *dent*, *gant*, *sang*, *Zan*, *rang*, and *lent*, consisting of an initial /p/, /t/, /k/, /b/, /d/, /g/, /s/, /z/, /r/, and /l/, respectively, followed by /ɑ̃/). Participants first made their choice by pressing an appropriate key. They then rated how well the syllable just presented (which they could hear again by pressing a key) matched the French keyword consonant they had indicated, using a 1–5 scale, in which 1="poor match" and 5="excellent match." The next trial was then administered until the test phase was completed.

B. Results

1. Discrimination

The performance of the French and Hebrew participants, pooled across vowels, is summarized in Figs. 1(a) and 1(b): correct discrimination and latencies for correct responses. Analyses of variance were conducted in two steps: first, only the structural factors of triad *Target* (dental versus velar-initial X in AXB trials) and triad *Pattern* (XXY vs XYY trials, usually referred to as primacy versus recency trials), as well as the *Vowel* context factor, were examined separately for Hebrew and French participants. The factors of critical interest, cluster *Liquid* (/r/ vs /l/) and cluster stop *Voicing* (voiced vs voiceless), were examined in further analyses of variance which included listeners' *Language* (French versus Hebrew) as a between-subject factor. Because the distributions of percent correct data often fail to meet the normality criterion, especially when they lie close to the upper boundary, we applied an arcsine transform to them in order to approximate more closely a normal distribution, and also ran analyses on the transformed data. The results of these analyses are re-

ported in the following only in cases of discrepancy with the analyses on raw percentages.
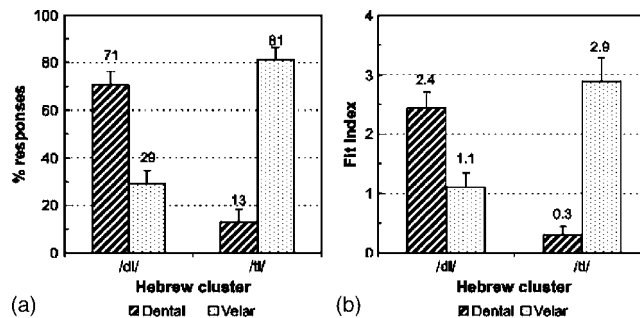
Triad Pattern had no significant effect in any analysis and will not be discussed further. Triad Target generally had a significant though numerically small effect, with slightly better discrimination performance for velar targets.[5] The triad Target effect did not otherwise interact with the main patterns found in the data, so we will not discuss it further.

In both the Hebrew and French percent correct discrimination and RT data, the Vowel factor had no significant main effect or interaction effects, which is not surprising for Hebrew participants, given their near-ceiling performance. Although Vowel had no significant effect overall, the French data showed numerical trends for poorer and slower discrimination in the -/li/ context (/d/-/g/ and /t/-/k/: 75.6 and 60.4% correct; 1408 and 1873 ms) than the -/la, lu/ contexts (/d/-/g/ and /t/-/k/: 77.8 and 65.7% correct; 1034 and 1420 ms).

The language comparison analyses for the percent correct data show that the Language × Liquid interaction was highly significant, $F(1,21)=128.8$, $p<0.00001$, reflecting that while French and Hebrew participants performed equally well for the /dr/-/gr/ and /tr/-/kr/ contrasts (above 97%), Hebrew participants outperformed French participants for the critical /dl/-/gl/ and /tl/-/kl/ contrasts (96 vs 71%), $F(1,21)=132.3$, $p<0.00001$. French performance was poorer for /tl/-/kl/ (64%) than for /dl/-/gl/ (77%), $F(1,11)=13.7$, $p<0.005$. For Hebrew participants, the differences across contrasts were numerically tiny. However, performance was significantly lower for the /tl/-/kl/ contrast (94.7%) than for the other contrasts pooled (97.9% on average), $F(1,10)=15.2$, $p<0.005$. This small difference remained significant in the arcsine-transformed data.

The RT data (Fig. 2) paralleled the percent correct data: the higher the discrimination score, the shorter the RT in each language group. The Hebrew listeners responded much more quickly overall than the French, $F(1,21)=58.50$, $p<0.00001$. Indeed, for the Hebrew listeners, the average response time, measured from the onset of the third stimulus, was negative (−916 ms), indicating that they actually responded without attending to the third stimulus. In contrast, the average RT for French participants was 977 ms, suggesting that, most of the time, they responded after they heard the third stimulus. In spite of this qualitative difference, a robust negative correlation between RT and percent correct

discrimination was found in both groups, showing that the association of "free" RT and percent correct data provides a coherent picture of discrimination difficulty. The correlations were computed on 12 pairs of data points: the dependent variable values (percent correct and RT) averaged across participants for each segmental contrast (4 onset-cluster contrasts × 3 vowels). For French participants, the correlation between percent correct score and RT was $r(10) = -0.96$, $p < 0.00001$. For Hebrew participants, although the range of variation in percent correct scores was very narrow (93%–99.7%), the correlation was nevertheless surprisingly robust, $r(10) = -0.79$, $p < 0.005$. The main trends in the percent correct data were thus found, in reverse, in the RT data: French participants were slower for the /l/ than for the /r/ cluster contrasts (1365 vs 590 ms), $F(1,11) = 27.0$, $p < 0.0005$, and slower for the /tl/-/kl/ than for the /dl/-/gl/ contrast (1571 vs 1158 ms), $F(1,11) = 17.5$, $p < 0.005$. For Hebrew listeners, the /tl/-/kl/ contrast yielded the slowest responses (by about 200 ms), $F(1,10) = 22.4$, $p < 0.001$, confirming the numerically small but reliable difficulty they encountered with this contrast.

2. Categorization

For all the stimuli, including /dl/ and /tl/, the rate of voicing confusion was negligible (below 1%). For the clusters that are legal in French (/gl, kl, gr, kr, dr, tr/), place confusions were rare (3.8% and 1.7% velar-to-dental for /gl, kl/ and /gr, kr/, 1.7% dental-to-labial confusions for /dr, tr/). For /dl, tl/, French participants reported mainly velar or dental responses; they seldom indicated labial responses (none for /dl/ items and 6% for /tl/ items). An analysis of variance was run with *Liquid* context (/l, r/), *Vowel* context (/a, i, u/), *Place* of articulation (dental, velar), and *Voicing* (voiced, voiceless) of stimulus initial stop as within subject factors. Percentage of "correct responses" was the dependent variable. A response was scored as correct if it was a stop with the stimulus' intended place of articulation, regardless of whether reported voicing was correct. Note that correct dental responses to /dl, tl/ items may correspond to either "faithful" perception of /dl, tl/ as a whole, or to perceptual repairs such as /dr, tr/, /t, d/, or /dəl, təl/. We treat all these as cases of faithful perception of the initial consonant. The Liquid × Place interaction was highly significant, $F(1,11) = 114.12$, $p < 0.00001$, reflecting nearly perfect performance for *all* the /r/ clusters (98.3%), regardless of Place but very different performance for the /kl, gl/ and /tl, dl/ clusters (96.2% vs 42%), $F(1,11) = 86.44$, $p < 0.00001$. In other words, perceptual shifts in place of articulation almost exclusively occurred for the /dl/ and /tl/ cluster stimuli. We therefore focused on the dental and velar place responses that were given to the /dl/ and /tl/ stimuli.

The raw percentages of dental, velar, and labial responses, according to cluster, are shown in Fig. 2(a). Following Guion *et al.* (2000), we also computed the "fit index" for each subject and each response type as the proportion of responses of a given type multiplied by the corresponding mean rating on a 1–5 scale. This produces fit indexes in the 0–5 range. The fit index data [Fig. 2(b)] yielded essentially the same patterns as the raw data. Analyses of variance were
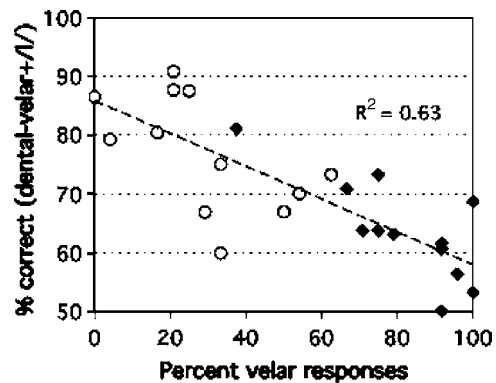
FIG. 3. French subjects: Correct discrimination of /dl/-/gl/ or /tl/-/kl/ as a function of the percentage of velar responses to /dl/ (open circles) or to /tl/ (closed diamonds), respectively.

run on the raw data (F_{raw}) and the fit index data (F_{fit}), with Voicing (/dl/ or /tl/) and Vowel (/a, i, u/) as within-subject factors. Response (dental or velar) was treated as a repeated measures factor. The Voicing × Response interaction was significant, (p's < 0.0001) reflecting a strong asymmetry in the perception of the /dl/ and /tl/. In /dl/, the initial consonant was less frequently judged to be velar than in /tl/, p's < 0.005, as can be seen in Fig. 2. This asymmetry is consistent with the better French performance in discriminating the /dl/-/gl/ contrast than the /tl/-/kl/ contrast. The discrimination performance correlated negatively with the rate of velar responses for the critical /dl, tl/ clusters, $r(22) = -0.79$, $p < 0.00001$ (Fig. 3: one data point per subject for each contrast type): the poorer the discrimination, the higher the velar response rate to the /dl, tl/ cluster involved.

Variation across vowel context was significant within the /dl/ and /tl/ stimuli, $F_{raw}(2,22) = 7.10$, $p < 0.005$; $F_{fit}(2,22) = 4.80$, $p < 0.05$. There were fewer velar responses with /a/ than with /i, u/, $F_{raw}(1,11) = 8.73$, $p < 0.05$; $F_{fit}(1,11) = 6.28$, $p < 0.05$ (/dl/ stimuli: 21% vs 29% and 38% velar responses with /a/ vs /i, u/; /tl/ stimuli: 67% vs 94% and 83%).

C. Discussion

Experiment 1 confirmed the robustness of the French listeners' dental-to-velar perceptual repair of utterance-initial /dl, tl/. The repair extends to non-native speech items beginning with /dl/ or /tl/ for French listeners, but does not extend to native listeners of the target language, Hebrew, in which initial /dl/ and /tl/ are permissible. Two important concerns about the interpretation Hallé *et al.* (1998) of the dental-to-velar shift as a case of "contextual perceptual assimilation" [Hallé *et al.* (1998), p. 604], have thus been resolved by the current study.

First, the phenomenon is not simply attributable to the intrinsic acoustic properties of the speech stimuli of Hallé *et al.* (1998), which could have reflected the French speaker's pronunciation difficulty in producing /dl/ and /tl/ targets. Monosyllabic items such as /tla/ or /dli/ produced by a native speaker of Hebrew, which allows word-initial /dl/ and /tl/, gave rise in the present experiment to a very clear pattern of dental-to-velar perceptual shift by French but not Hebrew listeners.

Second, the use of monosyllables and the cross-linguistic design, involving stimuli that have a clearly non-French accent, appear to have successfully minimized the likelihood of lexical influences (i.e., feedback) on the French listeners' performance. For example, the observed effects of vowel context were indeed inconsistent with lexical bias accounts of French listeners' performance. French listeners tended to have more difficulty discriminating /dl/-/gl/ and /tl/-/kl/ in the /i/ than in the /a/ or /u/ contexts and, in the categorization test, produced significantly more velar responses to /dl, tl/ in the /i/ or /u/ than in the /a/ context. This pattern indicates that dentals are the most confusable with velars to a French ear when they occur in the -/li/ context, and the least confusable in the -/la/ context. In the French lexicon, words beginning with /gli/ or /kli/ are outnumbered by those with /gla/ or /kla/ in terms of both types and tokens (see footnote 4) and therefore constitute smaller lexical cohorts. Thus, a cohort-based lexical bias would predict more dental-velar confusability for /dl/-/gl/ and /tl/-kl/ in the /a/ than in the /i/ context, a pattern which is exactly opposite to that observed. As for a lexical bias induced by individual words, only "clou," "glas," and "glou" could bias French listeners to hear a velar stop in /tlu/, /dla/, and /dlu/, predicting less dental-velar confusability in the /i/ than in the other vowel contexts, and more confusability in the /u/ context. Again, the variation in French performance according to vowel context, in both discrimination and categorization, is not consistent with the predictions based on such lexical bias. Nor is it explainable by acoustic differences: SCG differences would predict -/lu/ to be the most difficult context (according to Table II, dentals are acoustically more similar to velars in this context than they are to dentals in other contexts). Nor should dental-velar shifts be the most frequent in the -/li/ context, based on acoustic differences. One possible account of the stronger velar shift with /tli, dli/, however, is compensation for coarticulation, given that high-front /i/ presumably induces more anticipatory fronting than the other vowels.

To sum up, the dental-to-velar perceptual repair effect for /dl, tl/ is more likely induced by phonological factors (i.e., language-specific phonotactics) than by either low-level acoustic properties, coarticulatory properties, or lexical influences. If this is correct, the effect should be found for other languages which, like French, disallow /dl, tl/ word-initially. Yet, the asymmetry between /dl/ and /tl/ found in the previous studies on French, as well as in the discrimination data and even more dramatically in the categorization data of the present study, does not fit well with an exclusively phonotactic motivation. It suggests that the basic effect of phonotactic repair is modulated by factors other than a cluster's illegal status.

Those other factors could be phonetic. There is a better phonetic match between the French and Hebrew stops for /d/ than for /t/. The voiced stops of both Hebrew and French have clear prerelease voicing lead, but the Hebrew voiceless stops we used had a long lag aspirated VOT, unlike the typically short lag unaspirated voiceless stops in French. If the dental-to-velar perceptual repair effect is modulated by such language-specific phonetic factors in stop voicing settings,

rather than, for example, structural factors at a more abstract level (cf. Moreton, 2002), then native listeners of English should (1) show the dental-to-velar shift (i.e., phonotactic repair at the phonological level) but (2) should exhibit a voiced-voiceless /dl, tl/ asymmetry opposite to that in French listeners (i.e., language-specific phonetic modulation of the effect). This is because there is a better phonetic match between the English and Hebrew stops for /t/ than for /d/, contrary to the French case. The voicing distinction in English stops relies on the VOT difference between short lag and long lag, as there is not systematic prevoicing for English voiced stops (cf. Lisker and Abramson, 1964). We need, however, to first establish whether English listeners also perceive utterance-initial /dl, tl/ as velar, that is, display phonotactic repair at all.

In Experiment 2, we tested native listeners of English on their perception of Hebrew /tl, dl/. The first purpose was to establish whether English listeners display the same kind of perceptual shift as French listeners. If yes, this would strengthen the language-specific phonotactic repair account of this shift. To this date, indeed, *direct* evidence for a perceptual repair of /dl/ or /tl/ in languages that disallow them has been reported only for French. For English listeners, there is indirect evidence for a perceptual repair, showing that the location of the categorical boundary in /r/-/l/ continua is biased by a preceding consonant context toward legal obstruant-liquid responses (Massaro and Cohen, 1983; Pitt, 1998). If, as this indirect evidence suggests, English listeners also display a perceptual shift for /dl, tl/, a second purpose was to examine whether /dl/ and /tl/ are affected asymmetrically by the shift and if yes, whether the asymmetry is congruent with the English phonetic settings of the voicing distinction. Differences between French and English listeners in the perceptual asymmetry between /dl/ and /tl/, related to language differences in stop voicing settings, would provide evidence of language-specific *phonetic* modulation of the dental-to-velar perceptual repair.

III. EXPERIMENT 2

In this experiment, native speakers of American English, who lacked contact with any language allowing /dl, tl/ word-initially, were run on the same tasks and materials as the French participants in Experiment 1.

A. Method

1. Participants

Fourteen students at Wesleyan University (mean age 19 years, age range 18–22) participated in the experiment for course credit. All of them lacked exposure to Hebrew, or to any language allowing /dl, tl/ word-initially, and lacked experience with French. None of them reported any speaking, hearing, or reading problems. Five additional Americans were tested; four failed to complete all tests and one completed the tasks but was an outlier with respect to miss rate (3.4% versus average 0.25% for the retained subjects).

2906   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

P. A. Hallé and C. T. Best: Dental-to-velar perceptual assimilation
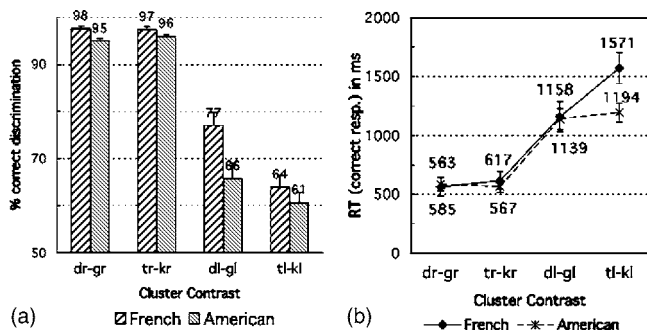
FIG. 4. Discrimination performance of American vs French subjects: (a) Percent correct discrimination, and (b) RT for correct responses in ms. Error bars indicate standard error.

2. Stimuli, design, and procedures

Stimuli were the same as in Experiment 1. The categorization test keywords were *pang*, *tang*, *kang*, *bang*, *dang*, *gang*, *sang*, *zang*, *rang*, and *lang*. All were words except *kang*, *lang*, and *zang*, which, however, all are possible and easy to pronounce surnames. They were modeled after the French keywords in Experiment 1: an initial /p, t, k, b, d, g, s, z, r, l/ followed by /æŋ/, an English rime that is phonologically similar to French /ɑ̃/.

B. Results

1. Discrimination

As can be seen in Fig. 4, the discrimination performance of the American participants was very similar to that of French participants. They performed close to ceiling level for the /dr/-/gr/ and /tr/-/kr/ contrasts, and much more poorly for the contrasts involving the illegal clusters /dl/ or /tl/ ($p < 0.00001$). They thus differed from Hebrew listeners in much the same way as French did. Analyses of variance similar to those run on the discrimination data in Experiment 1 were conducted, examining first the possible effect of the structural factors triad Pattern (primacy versus recency) and triad Target (dental versus velar) and Vowel context (/a, i, u/). As in Experiment 1, triad Pattern did not approach significance and will not be discussed further. Triad Target was significant—for the /l/ cluster contrasts only—with better performance for velar than dental targets (67% vs 59.5% correct), $F(1,13)=7.85$, $p < 0.05$. Vowel was not significant overall. However, Vowel and Liquid context interacted significantly, $F(2,26)=18.25$, $p < 0.0001$, reflecting inconsistent vowel effects for the /r/ cluster "baseline" contrasts and increasingly better discrimination performance from /i/ to /u/ to /a/ for the /l/ cluster contrasts (59%, 63%, and 68% correct, respectively): performance was the lowest for the /li/ context, $F(1,13)=7.10$, $p < 0.05$ (/li/ vs /lu, la/), and the highest for the /la/ context, $F(1,13)=17.38$, $p < 0.005$ (/la/ vs /lu, li/), just like in the French data. This pattern of performance is not consistent with the acoustic differences across vowels or with lexical influences for the same reasons as in the case of the French data. Again, as with the French listeners, the Americans' greater difficulty in the /i/ context could only be consistent with a compensation for coarticulation account.

Language comparison analyses were then run on the percent correct and RT data of all three listener groups, with Liquid context (/l/ vs /r/) and cluster stop Voicing (voiced versus voiceless) as within-subject factors and listener Language (American, French, and Hebrew) as a between-subject factor. American participants showed a pattern of results similar to that of French listeners but performed less well overall, $F(1,24)=8.25$, $p < 0.01$. Because Hebrew participants outperformed French participants for all contrasts, they obviously outperformed American participants as well.

We therefore focus on the French-American comparisons in the following. American participants performed less well than French participants on the /r/ clusters, $F(1,24)=12.19$, $p < 0.005$. They also performed less well on the /l/ clusters but the difference was significant only for /dl/-/gl/ (66% vs 77%), $F(1,24)=6.41$, $p < 0.05$, not for /tl/-/kl/ (61% vs 64%), $F(1,24)=1.01$, $p=0.33$. Whereas the French data showed a clear voicing asymmetry, with better performance on /dl/-/gl/ than /tl/-/kl/ (77% vs 64%, $p < 0.005$), the asymmetry was not clear-cut in the American data (66% vs 61%, $F(1,13)=3.2$, $p=0.092$), though also marginally in favor of the /dl/-/gl/ contrast. The weaker voicing asymmetry in the American than in the French data is reflected in a marginally significant Language × Voicing interaction for /l/ clusters alone, $F(1,24)=3.10$, $p=0.088$.

As in Experiment 1, the RT data paralleled the percent correct data. In the American data, RT negatively correlated with percent correct discrimination, $r(10)=-0.98$, $p < 0.00001$ (computed on 12 pairs of data points, as in Experiment 1). American participants were much slower for the /l/ than for the /r/ cluster contrasts (1167 vs 576 ms), $F(1,13)=38.73$, $p < 0.0001$. American participants, unlike French listeners, showed equivalent RTs for the /dl/-/gl/ and /tl/-/kl/ contrasts (1194 vs 1139 ms), $F < 1$. Hence, both percent correct and RT data indicate that the clear voiced-voiceless asymmetry found in the French discrimination data for the /l/ clusters was virtually absent in the American data. RTs for the /tl/-/kl/ contrast were shorter overall for American than for French participants (1194 vs 1571 ms) but the difference was statistically significant only after discarding two subjects (one French and one American) who had negative RTs, $t(22)=2.41$, $p < 0.05$; this contrast thus might have been somewhat easier for American than for French listeners with respect to response times.

2. Categorization

American participants showed virtually no place confusions for the legal clusters (/gl, kl, gr, kr, dr, tr/), except for one /f/ response to a /gli/ trial. For all the stimuli, including the critical /dl/ and /tl/ stimuli, the rate of voicing confusion was negligible (below 0.2%). Like the French participants, the American participants seldom gave labial responses for /dl/ or /tl/ (0% for /tl/ items, 0.6% for /dl/ items). An analysis of variance similar to that run on the French categorization data (Experiment 1) was run on the American categorization data, with the percentage of correct responses (see Experiment 1) as the dependent variable. However, Liquid context (/l/ vs /r/) could not be included as a factor because there was no variance for the subset of /r/ clusters (100% correct for
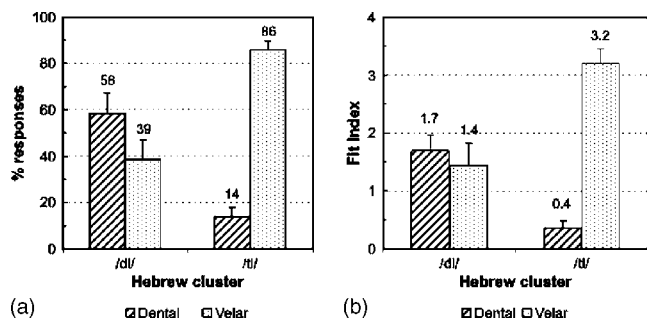
FIG. 5. Categorization of the initial stop of Hebrew /dl/ and /tl/ by American subjects: (a) Percentages of dental and velar responses, and (b) corresponding "fit indexes." Error bars indicate standard error.

FIG. 6. American subjects: Correct discrimination of /dl/-/gl/ or /tl/-/kl/ as a function of the percentage of velar responses to /dl/ (open circles) or to /tl/ (closed diamonds), respectively.

each cluster and each participant). The mean percentage of correct responses was 99.7% for /gl, kl/ whereas it was only 36% on average for /dl, tl/, $F(1,13)=123.99$, $p<0.00001$. We therefore focused on the dental and velar place responses that were given to the /dl/ and /tl/ stimuli, as in Experiment 1. The mean percentages of dental, velar, and labial responses according to cluster are shown in Fig. 5(a); the fit index data [Fig. 5(b)] exhibited essentially the same patterns. The American and French data were entered into two analyses of variance, one for the percentage data and the other for the fit index data, in which Language (American versus French) was a between-subject factor. As in Experiment 1, Voicing (/dl/ vs /tl/) and Vowel (/a, i, u/) were within-subject factors, and Response (dental versus velar) was treated as a repeated measures factor. The same voiced-voiceless asymmetry as in the French categorization data held for the American categorization results, reflected by a significant Voicing × Response interaction, p's <0.001. The initial consonant was less consistently judged as velar in /dl/ than it was in /tl/, p's <0.01. For the /dl/ stimuli, however, the advantage of dental over velar responses was less marked for American than French listeners, especially with respect to the fit index. The significant Language × Response interaction for the /dl/ stimuli in the fit index data, $F_{fit}(1,24)=5.34$, $p<0.05$, reflected the smaller fit index for dental responses to /dl/ stimuli in the American than the French data (1.7 vs 2.4), $F_{fit}(1,24)=10.09$, $p<0.005$. (This interaction was not significant in the raw percentage data.) There was thus some evidence for a smaller voicing asymmetry in the American than in the French categorization data for dental stop+/l/ stimuli. In a correlation analysis of the American data, similar to that run in Experiment 1, the percent correct score for /dl/-/gl/ and /tl/-/kl/ correlated negatively with the percentage of velar categorizations for /dl/ and /tl/: $r(26)=-0.62$, $p<0.0005$ (Fig. 6: one data point per subject for each contrast type, as in Fig. 3). The correlation was numerically smaller than that found for French listeners ($r(22)=-0.79$) but not significantly so (*Fisher Z-transformed difference*: -1.12, $p=0.26$).[6] That difference is nevertheless in line with the observation that discrimination data parallels identification data more closely for French than for American listeners: for American listeners, the voiced-voiceless asymmetry is significant in the identification data, not in the discrimination data; for French listeners, the asymmetry is
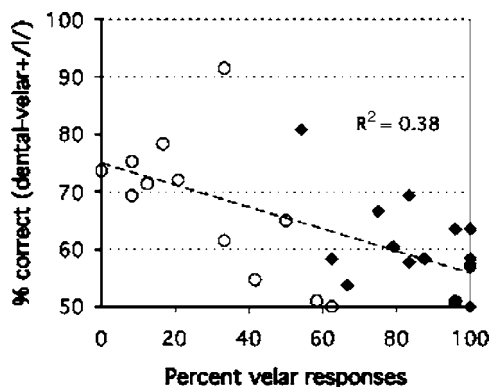
found for both and is stronger than that of Americans for the fit index of the identification data.

The American data were also similar to the French data with respect to the Vowel effect, with a significant variation of velar judgments across vowel context, $F_{raw}(2,26)=21.08$, $p<0.00001$; $F_{fit}(2,26)=5.44$, $p<0.05$. Velar judgments were less frequent with /a/ than with /i/ or /u/ overall (/dl/: 23%, 36%, and 57%, respectively; /tl/: 73%, 92%, and 92%).[7] This runs counter to the lexical bias account that would predict more velar judgments in the /a/ context, for essentially the same reason as for French participants.

C. Discussion

The discrimination test showed that American participants, just like French participants, had difficulty distinguishing /tl/ from /kl/ or /dl/ from /gl/. Thus, the basic phonotactic repair effect extends to other languages for which initial /dl, tl/ is illegal. Nonetheless, there are differences from the French pattern. American participants performed less well than French participants on all contrasts, and significantly so, except for /tl/-/kl/. For instance, they performed slightly but significantly less well on /dr/-/gr/ and /tr/-/kr/ (95.5% vs 97.5% correct). This could reflect a general perceptual difficulty induced by the Hebrew /r/ context, which is phonetically further away from English than French /r/. American English /r/ (bunched /r/ or retroflex /r/, cf. Boyce and Espy-Wilson, 1997; Zawadzky and Kuehn, 1980) is not uvular, as are both French and Hebrew /r/s (Delattre, 1969; Laufer, 1990). The same remark could apply to the lower performance overall of American participants for the /l/ contrasts: English /l/s tend to be "darker" ([ɫ]) than both French and Hebrew "light" /l/s. American listeners performed more poorly than French participants on /dl/-/gl/ (66% vs 77% correct discrimination) but not on /tl/-/kl/ (61% vs 64%), on which they were actually faster than French listeners. In the categorization test, they showed the same voicing asymmetry as did the French listeners, categorizing dentals as velars less often for /dl/ than for /tl/, although this pattern was less marked than in the French data. This particular pattern of similarities and differences between the American and French asymmetries clearly runs counter to the hypothesis that the language-specific phonetic settings of stop voicing in

P. A. Hallé and C. T. Best: Dental-to-velar perceptual assimilation

French versus English affect the /dl, tl/ → /gl, kl/ phonotactic repair effect. This hypothesis predicted that the voiced-voiceless asymmetry in the effect would go in opposite directions for English and French listeners. Although the lower discrimination performance of American than French listeners on the voiced contrast /dl/-/gl/ could reflect a real difference in the way French and American listeners perceive language-specific phonetic specifications for voiced stops, we found a similar asymmetry in the repair effect for both groups in the categorization data. The *direction* of asymmetry in phonotactic repair thus is not driven by the specific phonetic settings of stop voicing of the listeners' native language, although those details do appear to modulate its *magnitude*. Some other factor must be responsible for the stronger dental-to-velar shift for /tl/ than for /dl/, observed indeed even for Hebrew listeners to a very small but significant extent; we take up this issue in Sec. IV.

IV. GENERAL DISCUSSION

In Experiment 1, French listeners had substantial difficulty discriminating the Hebrew dental-velar stop +/l/ contrasts, especially /tl/-/kl/. Consistent with their discrimination performance, they often categorized Hebrew /tl/ and /dl/ as velar-initial, especially /tl/. They thus showed a strong dental-to-velar shift with Hebrew /tl/ and, to a lesser extent, /dl/. Their performance was contrary to any account in terms of lexical bias, suggesting that the dental-to-velar shift operates at a sublexical level of perception. Surprisingly, Hebrew listeners had some difficulty discriminating /tl/-/kl/, although they performed near ceiling. Experiment 2 extended the findings to native speakers of American English, who differed from French listeners in their equally low discrimination performance on /dl/-/gl/ and /tl/-/kl/, and showed somewhat less /dl, tl/ asymmetry in their categorization responses.

The results provide unequivocal answers to several key issues raised in Sec. I. First, using /dl, tl/ stimuli produced by a native speaker of Hebrew, hence stimuli that must convey a dental quality because both /dl/-/gl/ and /tl/-/kl/ are contrastive in Hebrew, did not prevent French and American listeners from experiencing a dental-to-velar perceptual shift, as in the study of Hallé *et al.* (1998), which used stimuli produced by a native speaker of French. Second, the dental-to-velar shift was not modulated by lexical feedback for either the French or the American listeners, and likely operated at a sublexical level of speech perception. The results also provide a clear answer to the main issue that was raised: The dental-to-velar shift effect is language-specific rather than universal. Indeed, the effect is substantial in the two languages examined that disallow /dl/ and /tl/ in initial position, but in Hebrew, which allows initial /dl/ and /tl/, the effect is either absent altogether (/dl/-/gl/) or numerically tiny and still near-ceiling (/tl/-/kl/). A claim that the dental-to-velar shift is due to a *universal* perceptual bias toward hearing a dorsal stop in coronal stop + coronal liquid clusters is thus untenable. For example, the universal compensation for coarticulation mechanism suggested by Dupoux *et al.* (2001) as a possible explanation of the dental-to-velar shift obviously fails to account for the large difference between Hebrew and

French or American data. Yet, as the slight difficulty encountered by Hebrew listeners with /tl/-/kl/ may suggest, we should ask whether there remains some residue of experience-independent perceptual difficulty with word-initial coronal-dorsal stop +/l/ contrasts. To probe this issue, we turn to the question of universal tendencies across languages.

A. Universal tendencies

Diachronic and synchronic data indicate that /tl/-/kl/ and /dl/-/kl/ are difficult contrasts. Diachronically, coronal-dorsal contrasts in stop +/l/ clusters have often become neutralized in the languages in which they have existed (cf. Hallé *et al.*, 1998, for a brief survey), suggesting intrinsic difficulty with these contrasts. Importantly, though, in those languages where a coronal-dorsal distinction in /l/ clusters is lacking or has been lost, it is often the coronal, not the dorsal clusters that were maintained. This holds for languages of quite different linguistic families (Hmong-Mien dialects, Setswana or Sesotho in the Bantu family, Aztec languages such as Nahuatl, etc.), suggesting that the difficulty does not lie in the low perceptibility of the coronal stop +/l/ clusters per se, but, rather, in the auditory similarity of the coronal and dorsal stop +/l/ clusters. It is also true that coronal stop +/l/ clusters are unstable diachronically and have a much lower incidence than dorsal stop +/l/ clusters in the languages of the world, a fact which would usually be interpreted as reflecting a structural constraint against adjacent coronal consonants: avoidance of homorganic CC sequences, especially when the two Cs are close in sonority (Padgett, 1991; Selkirk, 1988). Avoidance of coronal stop +/l/ clusters may alternatively be understood as resulting from the avoidance of clusters that are confusable with the readily acceptable dorsal stop +/l/ clusters.

Bradley (2006) proposed that coronal stop +/l/ clusters are intrinsically confusable with their dorsal counterparts, due to the articulatory adjustments necessary to produce them. Bradley reasoned that "In oral stops that are released into a following lateral, coronal and velar constrictions are produced at or behind the lateral constriction. As a result, these stops are not well differentiated by their release bursts …" Flemming (2002) proposes that clusters such as [tl] and [kl] share the same critical "auditory features" ($F2$ transition, burst diffuseness and frequency) and are thus poorly discernible, whereas they both differ from [pl] on all three features. Our acoustic measurements of the Hebrew dorsal and coronal stop +liquid clusters (Table II) also suggest that the bursts of /tl/ and /kl/ (or /dl/ and /gl/) are less well differentiated than those of /tr/ and /kr/ (or /dr/ and /gr/).

It is difficult, however, to conclude that the physical (or psychophysical) similarity between coronal and dorsal stops released into /l/ is sufficient to universally entail perceptual confusion. The much better discrimination performance of Hebrew listeners compared to French or American listeners instead suggests that a phonological grammar allowing /tl/-/kl/ (and possibly /dl/-/gl/) greatly enhances this otherwise poorly perceptible contrast: Hebrew listeners are tuned to attend to the subtle acoustic-articulatory differences between /tl/ and /kl/, however small these differences are. Because

French and English both disallow /tl, dl/ while allowing /kl, gl/, the discrimination data of American and French listeners could provide measures of objective (i.e., unbiased by native phonological grammar) similarity between /tl/ and /kl/, and between /dl/ and /gl/. French and American listeners indeed exhibited roughly the same dental-to-velar perceptual shift for /dl, tl/. Yet, systematic differences in magnitude for the /dl, tl/ asymmetry, in both their discrimination and identification data, suggest their performance is influenced by language-specific aspects other than purely phonological ones.

B. Differences between French and English

The difference between the French and American data is not dramatic but, still, is robust. While the discrimination performance of French and American listeners was virtually the same on /tl/-/kl/, French listeners outperformed American listeners on /dl/-/gl/, and showed more marked /dl, tl/ asymmetry in their categorization data. Because French and English equally disallow /dl, tl/, these differences must be explained by some language-specific factors other than the phonological grammar alone.

One such factor is the language-specific phonetic settings for stop voicing. The better match of Hebrew /d/ ([d̪]) with French /d/ ([d̪]) than English /d/ ([t/d]) and, conversely, the better match of Hebrew /t/ ([t̪ʰ]) with English /t/ ([tʰ]) than French /t/ ([t̪]), could have led to opposite patterns of /dl, tl/ asymmetry for French and American listeners. But this did not occur. The two groups of listeners differed in the magnitude of the /dl, tl/ asymmetry, not in its direction. Both groups exhibited more dental-to-velar shifts for /tl/ than for /dl/ in their identification data and, congruent with identification, lower discrimination performance for /tl/-/kl/ than /dl/-/gl/ (at least, numerically). The French data differed from the American data in that the advantage for /dl/-/gl/ over /tl/-/kl/ was substantial only in the French data, not in the American data. The pattern of results is perhaps best captured in the following way: both French and American listeners performed poorly on the /tl/-/kl/ contrast—slightly above 60% correct discrimination—but only French and not American listeners exhibited improved performance for /dl/-/gl/.

The reason why American listeners did not perform better on /dl/-/gl/ than /tl/-/kl/ could be their lack of consistent experience with voicing leads in phonetically voiced stops. They might be deaf to these voicing leads. French listeners, in contrast, are likely to perceive the voicing murmur of pre-voiced [d]s, which is the main cue to the voicing contrast in French stops. They thus could parse inputs such as murmur+[d]+[l]+[a] into acceptable disyllabic sequences close to /əd.la/.[8] American listeners would not be able to produce that parse because they just do not hear the murmur. Another possibility is that French but not American listeners perceive and use the place information that could be conveyed by voicing leads. Although we are not aware of acoustic or perceptual studies demonstrating that cues to place of articulation are present in voicing leads and are exploited in perception, this possibility deserves consideration. While

prerelease voicing murmurs are low in amplitude, they are somewhat audible and could contain place information insofar as the articulators arrive at their intended place of articulation during stop closure. We did measure spectral differences between the voicing leads of the Hebrew dental and velar stops stimuli used, and found some differences: "velar" murmurs tend to have a lower $F1$ and a higher $F2$ than "dental" ones. Yet, further investigation is needed to determine whether place information in voicing leads is perceptible, at least to listeners whose native language employs voicing lead.

To sum up, while the main determinants of the dental-to-velar perceptual repair effect must lie at an abstract phonological level (the native phonological grammar), we nevertheless found some modulation of the effect by the language-specific phonetic settings of voicing in stops. French listeners perceived Hebrew /dl/ more "faithfully"—as a coronal-initial cluster—than American listeners did. Yet, common to both groups of listeners, Hebrew /dl/ induced less dental-to-velar shift than /tl/, and this is most clearly evidenced in the categorization data. We dubbed this common pattern "voicing asymmetry." We turn to this pervading aspect of the data, found throughout the present study as well as in our previous work.

C. The /dl, tl/ asymmetry

Voicing asymmetry, which suggests higher perceptibility of dental place in Hebrew /dl/s than /tl/s, was found for all three groups but to different extents. It was especially clear in the categorization data of French and American listeners, though somewhat less so in the latter group. As for discrimination, the asymmetry was substantial in the French data, very tiny and near ceiling in the Hebrew data, and nonsignificant in the American data despite numerical trends also suggesting easier discrimination for /dl/-/gl/ than /tl/-/kl/.

Intriguingly, the voicing asymmetry we consistently observed is at odds with the distribution of the /dl/-/gl/ and /tl/-/kl/ contrasts in the languages of the world. While both contrasts are infrequent across languages, /dl/-/gl/ is even less frequent than /tl/-/kl/, as shown by cross-linguistic evidence: if a language allows /dl/-/gl/, it also allows /tl/-/kl/ but not vice-versa [cf. Tobin's (2002) survey of forty languages with complex onset clusters].[9] Assuming that languages avoid contrasts doomed by perceptual similarity, the cross-linguistic pattern would suggest that /dl/-/gl/ is less perceptible than /tl/-/kl/, rather than the other way around.

In the account offered by Flemming (2002; also see Bradley, 2006), the lesser stability across languages of /dl/-/gl/ than /tl/-/kl/ is attributable to the poorer acoustic information appearing in the release burst of voiced than voiceless stops. There is generally less information in the bursts of /dl, gl/ than /tl, kl/ to distinguish velars from dentals because voiced stops have quieter bursts than voiceless stops (Zue, 1976). Our measurements of burst integrated energy in the Hebrew clusters (Table II) indeed confirm Zue's observation, with the qualification that the velar-dental difference,

in terms of burst energy, is smaller for voiced than voiceless stops, thus predicting better discrimination for /tl/-/kl/ than for /dl/-/gl/.

The "quiet burst" argument, as well as the cross-linguistic data, are thus at odds with the asymmetry we observed. A possible explanation is that stop voicing contrasts generally involve a short versus long lag VOT voicing distinction, not a prevoiced versus short or medium lag as in French or Hebrew. There may be perceivable information on place of articulation in the voicing lead of /dl/ and /gl/, as we discussed earlier, but this remains quite speculative at the moment. The /dl, tl/ asymmetry is also in line with the larger velar-dental SCG differential for /dl/-/gl/ than for /tl/-/kl/ (250 vs 100 Hz), but this SCG differential varied across vowel context and did not reach significance. Finally, no discernible acoustic difference was found between velar- and dental-initial /l/ clusters in either the /l/ or the following vowel formant patterns, whether the initial consonant was voiced or not. The formant patterns thus cannot explain the observed advantage of /dl/-/gl/ over /tl/-/kl/ in discrimination. Overall, then, the acoustic evidence does not explain well why /dl/-/gl/ is discriminated more easily than /tl/-/kl/: /tl/ does not clearly appear to be physically (or psychophysically) closer to /kl/ than /dl/ is to /gl/ in our Hebrew stimuli.

What about, as an alternative, "structural" linguistic arguments? Is the greater acceptability of /dl/ explainable by structural constraints? An example of differential structural acceptability has been recently offered by Moreton (2002), who showed there is a stronger perceptual bias against utterance-initial /dl/ than /bw/ for English-speaking listeners, although both clusters are illegal in English. Moreton interpreted this finding in terms of universal structural constraints, reasoning that "the *[dl] sequence is closer in sonority than *[bw]…, and hence a worse structural violation" [Moreton (2002), p. 57]. Indeed, there is a smaller differential in sonority in /dl/ than in /bw/ because /l/ is less sonorous than /w/ in proposed sonority scales (Clements, 1990; Kahn, 1980). The recent work of Berent *et al.* (in press) also demonstrates the role of sonority profile in cluster acceptability. But the problem in our case is that the differential sonority account predicts that /dl/ is just as unacceptable as /tl/ because the stop and the liquid are not further apart in sonority in /dl/ than in /tl/. Indeed, they might even be closer in /dl/ than /tl/, as voiced /d/ might possibly be considered more sonorous than voiceless /t/. Yet, the lesser acceptability of /tl/ is resistant to language-specific phonetic differences. It holds for speakers of French, English, and even Hebrew, suggesting that the asymmetry is driven by some structural, universal factor. Further research will be needed to pinpoint what kind of structural factor could adequately account for the lesser acceptability, hence the greater dental-to-velar perceptual shift, for /tl/ than /dl/.

D. Models of cross-language speech perception

Theoretically speaking, it is also important to ask whether extant models of non-native speech perception can contribute to our understanding of the dental-to-velar perceptual shift. For this discussion, we limit ourselves to three widely known models: Speech Learning Model (SLM: Flege, 1995), Native Language Magnet (NLM: Kuhl and Iverson, 1995), and Perceptual Assimilation Model (PAM: Best, 1995). Could any of these models predict a dental-to-velar perceptual shift repairing /tl, dl/ into /kl, gl/?[10] How would the models explain the stronger repair effect for /tl/ than /dl/, for both English and French native listeners?

A general point must be made first, concerning the issue of whether stop + liquid clusters should be viewed analytically as a sequence of two phonemes, or rather, as unanalyzed syllable onsets. In the former case, all three models would agree that the perception of /dl, tl/ is reducible to that of /d, t/ plus /l/, and none would easily address the possible influence of the following /l/. In the second case only, if the clusters are perceived holistically as syllable onsets, might the three models differ in their predictions.

Although none of the three models explicitly addresses the issue of complex onsets, it is reasonable to assume that all should base their predictions about complex onset perception on *perceptual* distances. Such distances are conceivably determined both by objective stimulus properties and by how listeners represent speech sounds, depending on their native language. Models may differ on the latter issue of internal representations: NLM regards native categories as emerging from statistical clustering of auditory cues, whereas PAM instead regards them as the natural product of perceiving articulatory gestures that correspond to natively produced gestures and as organized in a phonologically principled way. SLM is neutral with regards to this issue. As for the "objective" component in perceptual distances, both NLM and SLM posit they are determined by the auditory properties of the stimuli, whereas PAM posits they are instead determined by their articulatory properties.

As discussed earlier, coronal stops released into /l/ may be intrinsically confusable with dorsal but not with labial stops for acoustic-articulatory reasons: dorsal and coronal but not labial constrictions are masked by the lateral constriction gesture (Bradley, 2006; Kawasaki, 1982). This distinction between nonlabial and labial stop +/l/ would be captured in Articulatory Phonology by the major within- versus between-organ tier distinction (Goldstein *et al.*, 2006; Goldstein and Fowler, 2003): /tl/-/kl/ is a within-tier distinction (tongue tip versus tongue dorsum, both on the tongue tier) and /tl/-/pl/ a between-tier distinction (tongue versus lips), which is presumably easier (Goldstein and Fowler, 2003). PAM would follow Articulatory Phonology to account for the perception of /dl, tl/, whereas NLM and SLM would reason in terms of psychophysical distances. But all three models would agree that /gl, kl/ is a closer match with /dl, tl/ than /bl, pl/. Would they also agree that dental-to-velar shift is more likely with /dl/ than /tl/? NLM and SLM, as they reason in terms of psychophysical distances, would have difficulty explaining the asymmetry because the evidence for larger psychophysical distances between /dl/ and /gl/ than between /tl/ and /kl/ remains, at the moment, quite weak (SCG differentials) or speculative (perception of the prevoicing murmur), and would have to address other acoustic measures that could favor the voiceless stops over the voiced

ones (e.g., burst intensity). PAM, as it reasons in terms of underlying gestural dynamics, might offer a more principled account of the asymmetry.

The particular formalization used by the models does not help, either, to set them apart on the issue of voicing asymmetry. For SLM, non-native speech sounds that are "similar" to some native category cause production and perception difficulties, due to interference from the native language; "new" speech sounds are easier because they escape such interference. Based on the observed data, that is, on a post hoc analysis, SLM could formalize the voicing asymmetry in treating Hebrew /tl/, which is massively perceived as /kl/ by French and American listeners, as "similar" to a native /kl/, and /dl/, which receives mitigated judgments, as a "new" speech sound. For NLM, discrimination is more difficult around prototypes than nonprototypes. The better discrimination of /dl/-/gl/ than /tl/-/kl/ could be taken, then, to indicate lower typicality of Hebrew /dl/ as an exemplar of French or English /gl/ than that of /tl/ as an exemplar of /kl/. Again, this is a post hoc interpretation. PAM could also offer a post hoc interpretation of the data based on the majority of velar judgments for /tl/ and of dental judgments for /dl/, considering /tl/-/kl/ as a single-category assimilation, and /dl/-/gl/ as a category-goodness assimilation type, an easier type of contrast in PAM's classification. This interpretation would be just as problematic as that of SLM and NLM if PAM also based its predictions on psychophysical distances and auditory representations. But PAM explicitly posits that speech perception is driven by both the native phonological grammar and by the detection of underlying articulatory gesture properties. A plausible account from the PAM perspective might thus be framed in terms of gestural organization.

E. A speculative articulatory account

So far, there is no definitively satisfying answer to the question of why /dl/ is more discernible from /gl/ than /tl/ is from /kl/, and more easily categorized as dental than /tl/. The answer to the puzzle might lie in the different gestural organizations that could apply to the production of /tl/ and /dl/. For all stop +/l/ onsets, the production of the lateral segment /l/ requires retraction of the side(s) of the tongue, thinning it laterally by anterior extension of the tongue blade or tongue tip (TT), along with posterior movement of the tongue dorsum (TD). This maneuver, then, involves constriction at both the TT and TD. These constrictions are relatively innocuous to the production of a velar stop +/l/ onset, in which the stop involves a TD constriction anyway, but are somewhat detrimental to the dental quality of intended dental stop +/l/ onsets. We speculate that the difference between /dl/ and /tl/ may reflect different intergestural phasing relationships between the stop and the /l/ for the voiced versus voiceless stops, due to the fact that /tl/ requires a glottal abduction gesture during /t/, whereas this glottal gesture is absent for both /l/ and /d/. In other words, the phasing of /d/+/l/ is gesturally simpler and therefore less constrained, because there are fewer gestures to phase to one another. The perceptual side of this tentative gestural dynamics account is that

(1) the intended dental gesture in /dl, tl/ may be misperceived as dorsal because it is contaminated by the tongue dorsum movement and (2) the confusion is all the more likely for the tighter gestural phasing relationship involved in [tl] than in [dl]. Thus, factoring out the intended place of articulation in the occlusion is conceivably more difficult in [tl] than in [dl]. This gestural dynamics account, however speculative at the moment, has the merit of (1) providing a low-level articulatory motivation of the dental-to-velar shift in general, and (2) accounting for the asymmetry in the perception of /dl/ versus /tl/. Unfortunately, there are no published studies on the gestural organization of English, French, or Hebrew stop +liquid onset clusters. The speculations offered here thus clearly call for further research.

Regardless of whether speech perception is determined by auditory analysis of acoustic events or recovery of articulatory gestures, it is clear from our findings that non-native speech perception is systematically biased not only by individual native segments and contrasts, but also by phonotactic constraints. Perception of segmental contrasts that do exist in the native language is disturbed when the segments involved are part of a phonotactic pattern that is not permitted by that language. All extant models of non-native speech perception fall short of anticipating these multisegmental perceptual difficulties, because they have thus far focused only on perception of singleton segments. If we attempt to extend them to account for onset clusters, all would presumably agree that /kl/ is the best repair for [*]/tl/ based on either acoustic or articulatory grounds. Yet, one intriguing pattern we consistently found, the /dl, tl/ asymmetry, is perhaps more easily explained in terms of articulatory gesture organization, in line with the current views of Articulatory Phonology. We provisionally favor PAM because it implicitly adopts the Articulatory Phonology framework. However, further articulatory studies of gestural organization in stop +liquid clusters are needed for a better understanding of their perception.

To summarize, much work is needed to develop predictions regarding multisegmental influences (such as complex syllable onsets). More research is needed, as well, on understanding the relationships between the articulatory organization of speech at the relevant phonetic and phonological levels, and the perception of native and non-native speech patterns, especially cross-linguistically.

[1]In the context of loanword phonology, words borrowed from a source language are introduced in the borrowing language by speakers who are aware of the necessary adjustments for the loanword to comply with the target language phonology (see, for example, Paradis and LaCharité, 1997). Loanword phonology primarily describes how "input" forms from a source language are adapted into "output" forms in a target language. In the context of speech perception, "phonological repair" refers to repairs made at a perceptual level by naive listeners who may not be aware of a phonological violation. Hence, the "perceptual repairs" we refer to do not involve conscious computations or cognitive strategies but instead occur automatically. Yet, the difference in meaning between loanword adaptation and perceptual repairs is rhetorical if one adheres to the view that adaptations largely reflect the layperson's perceptual assimilations (Vendelin and Peperkamp, 2004).

[2]In the languages which use vowel epenthesis in loanword adaptations, the inserted vowel is typically the least salient one (closest to ∅), that is, the phonetically most unmarked vowel in the vowel system (Kenstowicz, 2003). In Japanese, the default epenthetic vowel is /u/ realized [ɯ], the shortest vowel in Japanese (Han, 1962); only /u/ and /i/ undergo devoicing or reduction, and /u/ is the more central of the two (Keating and Huffman, 1984). The case of *drama* adapted into *dorama* or *zurama* (/dzurama/) is somewhat more complex. Because Japanese lacks /du/ in its /d/-initial syllables and has /dzu/ instead, a second repairing step is necessary with two possible solutions: maintain /d/ and use the nondefault /o/ vowel, or use the default /u/ vowel and substitute /dz/ for /d/.

[3]Reports differ as to whether trilled /r/s are often rather than occasionally produced by Hebrew speakers. Trilling would be more frequent in the Oriental variety of Hebrew, but tends to spread to the non-Oriental variety as well (Laufer, 1990). Trilled /r/s occur in some varieties of continental French. They are rare in current Parisian pronunciation, but do occur in restricted circumstances (e.g., in songs), and should not be unfamiliar to French listeners.

[4]In French, the number of types is 46, 28, and 7 for /kla/, /kli/, and /klu/ words, respectively; the number of tokens is 1634, 940, and 353. For /gl/ words, the number of types is 23, 13, and 10 for /gla/, /gli/, and /glu/, respectively; the number of tokens is 1108, 584, and 271. These counts are drawn from the Brulex database (Content *et al.*, 1990) based on a 23.5 million word corpus. In English, the number of types is 46, 12, and 2 for /kla/, /kli/, and /klu/ words, respectively; the number of tokens is 174, 90, and 15. For /gl/ words, the number of types is 29, 9, and 11 for /gla/, /gli/, and /glu/, respectively; the number of tokens is 205, 12, and 40 (from the one million words Brown corpus: Francis and Kucera, 1982).

[5]In the Hebrew participants' data, the advantage of velar over dental targets was significant, although numerically tiny, for the percent correct but not for the RT discrimination performance. In the French data, the advantage was found for both percent correct and RT performance but was confined to the /l/ cluster contrasts (75.1% vs 69% correct; 866 vs 1281 ms) and virtually absent in the /r/ cluster contrasts (97.7% vs 97.2%; 284 vs 337 ms).

[6]The following function (`compcorr`) can be used in the R environment (Becker *et al.*, 1988) to test for difference in correlation strengths (independent samples):

```
compcorr <- function(n1, r1, n2, r2)
{
# compares two correlation coefficients; returns difference
and p-value
# Fisher Z-transform
zf1 <- 0.5*log((1 + r1)/(1 - r1))
zf2 <- 0.5*log ((1 + r2)/(1 - r2))
# difference and p-value returned as list (diff, pval)
dz <- (zf1 - zf2)/sqrt(1/(n1 - 3) + (1/(n2 - 3)))
pv <- 2*(1 - pnorm(abs(dz)))
return (list(diff=dz, pval=pv))
}
```

[7]The lesser incidence of velar judgments for /dl, tl/ in the /a/ context is consistent with the higher discriminability of /dl/-/gl/ or /tl/-/kl/ in that context for both American and French listeners. No such correlation is observed in the case of /dlu, tlu/, which should be more confusable with /glu, klu/ on acoustic grounds but actually induced less dental-to-velar

shifts than /dli, tli/, or in the case of /dli, tli/, which led to the largest amount of velar shifts but should not have done so, given their acoustic characteristics, as suggested by Table II.

[8]Hallé *et al.* (1998) proposed that the voicing murmur of French [d], a "weakly resonant vowel-like sound" (Hallé *et al.* (1998), p. 605) could be perceived as a vowel and enhance the perceptibility of /dl/ due to the fact that, in French, dental stop +/l/ clusters are acceptable after a vowel (word-medially) because they allow for a phonotactically legal parse in this position, with a syllable boundary before /l/ (Dell, 1995; for a review, see Goslin and Frauenfelder, 2000).

[9]This cross-linguistic pattern also holds for languages in which voiced stops are phonetically prevoiced, such as Mexican Spanish. Mexican Spanish has few /tl/-words, most of them borrowed from Nahuatl (e.g., *tlaco* "coin," *tlapaleria* "hardware store") but no /dl/-word. One exception is Creole French (both the Indian Ocean and Caribbean varieties), in which word-initial /dl/ but *not* /tl/ is attested. Word-initial /dl/ appears in adaptations from French words in which agglutination is involved, such as [dle] for "lait" (from *du lait*, "some milk"), or quite commonly, [dlo] for "eau" (from *de l'eau*, "some water") as in *mon dlo lé au fé* ("my water is on the stove"). These forms alternate with [dəle] or [dəlo] (Chaudenson, 1974, p. 654).

[10]Other possible repairs of, for example, /tla/ are simplification (/ta/), epenthetic vowel insertion (/təla/), liquid substitution (/tra/), and labial substitution (/pla/). Labial responses suggesting labial substitution were negligible in our identification data. Dental responses to, for example, /tla/ might have occasionally corresponded to /ta/, /təla/, or /tra/ repairs. These repairs, however, very rarely appeared in open response transcription tasks (Hallé *et al.*, 1998) and are not compatible with the difficulties encountered by non-Hebrew listeners with /dl/-/gl/ or /tl/-/kl/. We therefore conservatively restrain the discussion to velar substitution repairs.

Becker, R., Chambers, I., and Wilks, A. (**1988**). *The New S Language* (Chapman & Hall, London).

Berent, I., Steriade, D., Lennertz, T., and Vaknin, V. (in press **2007**). "What we know about what we never heard: Evidence from perceptual illusions," Cognition.

Best, C. (**1995**). "A direct realist perspective on cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange and J. Jenkins (York, Timonium, MD), pp. 171–204.

Blumstein, S., and Stevens, K. (**1980**). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**, 648–662.

Boyce, S., and Espy-Wilson, C. (**1997**). "Coarticulatory stability in American English /r/," J. Acoust. Soc. Am. **101**, 3741–3753.

Bradley, T. G. (**2006**). "Contrast and markedness in complex onset phonotactics," Southwest J. Linguistics **25**, 29–58.

Chafcouloff, M. (**1979**). "Les propriétés acoustiques de /j, ɥ, w, r, l/ en français," ("Acoustic characteristics of French /j, ɥ, w, r, l"), Travaux de l'Institut de Phonétique d'Aix, Rep. No. 6, pp. 11–24.

Chaudenson, R. (**1974**). *Le Lexique du Parler Créole* (Lexicon of Creole, Honoré Champion, Paris).

Chayen, M. J. (**1973**). *The Phonetics of Modern Hebrew* (Mouton, Paris).

Cho, T.-H., and Ladefoged, P. (**1999**). "Variation and universals in VOT: Evidence from 18 languages," J. Phonetics **27**, 207–229.

Clements, G. N. (**1990**). "The role of the sonority cycle in core syllabification," in *Between the Grammar and the Physics of Speech*, edited by J. Kingston and M. Beckman (Cambridge University Press, New York), pp. 283–333.

Content, A., Mousty, P., and Radeau, M. (**1990**). "Une base de données lexicales informatisée pour le français écrit et parlé," ("A computerized lexical database for written and spoken French"), Annee Psychol. **90**, 551–566.

Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (**2000**). "Electrophysiological correlates of phonological processing: A cross-linguistic study," J. Cogn Neurosci. **12**, 635–647.

Delattre, P. (**1969**). "L' R parisien et autres sons du pharynx," ("Parisian R and other pharyngeal speech sounds"),The French Review **43**, 5–22.

Dell, F. (**1995**). "Consonant clusters and phonological syllables in French," Lingua **95**, 5–26.

Devens, M. (**1978**). "The phonetics of Israeli Hebrew: 'Oriental' versus 'General' Israeli Hebrew," Ph.D. dissertation, UCLA, Los Angeles, CA.

Devens, M. (**1980**). "Oriental Israeli Hebrew: A study in phonetics," Afroasiatic Linguistics **7**, 127–142.

Dupoux, E., Kakehi, Y., Hirose, C., Pallier, C., and Mehler, J. (**1999**). "Ep-

enthetic vowels in Japanese: A perceptual illusion?" J. Exp. Psychol. Hum. Percept. Perform. **25**, 1568–1578.

Dupoux, E., Pallier, C., Kakehi, Y., and Mehler, J. (**2001**). "New evidence for prelexical phonological processing in word recognition," Lang. Cognit. Processes **16**, 491–505.

Fischer-Jørgensen, E. (**1954**). "Acoustic analysis of stop consonants," Miscellanea Phonetica **2**, 42–59.

Flege, J. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange and J. Jenkins (York, Timonium, MD), pp. 233–277.

Flemming, E. (**2002**). *Auditory Representations in Phonology* (Routledge, New York).

Fowler, C., Brown, J., Sabadini, L., and Weihing, J. (**2003**). "Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks," J. Mem. Lang. **49**, 396–413.

Francis, W., and Kucera, H. (**1982**). *Frequency Analysis of English Usage: Lexicon and Grammar* (Houghton-Mifflin, Boston).

Goldstein, L., Byrd, D., and Saltzman, E. (**2006**). "The role of vocal tract gestural action units in understanding the evolution of phonology," in *From Action to Language: The Mirror Neuron System*, edited by M. Arbib (Cambridge University Press, Cambridge, MA), pp. 215–249.

Goldstein, L., and Fowler, C. (**2003**). "Articulatory phonology: A phonology for public language use," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin), pp. 159–208.

Goslin, J., and Frauenfelder, U. (**2000**). "A comparison of theoretical and human syllabification," Lang Speech **44**, 409–436.

Guion, S., Flege, J., Akahane-Yamada, R., and Pruitt, J. (**2000**). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**, 2711–2724.

Halle, M., Hugues, G., and Radley, J.-P. (**1957**). "Acoustic properties of stop consonants," J. Acoust. Soc. Am. **29**, 107–116.

Hallé, P., Best, C., and Levitt, A. (**1999**). "Phonetic vs. phonological influences on French listeners' perception of American English approximants," J. Phonetics **27**, 281–306.

Hallé, P., Chang, Y.-C., and Best, C. (**2004**). "Categorical perception of Taiwan Mandarin Chinese tones by Chinese versus French native speakers," J. Phonetics **32**, 395–421.

Hallé, P., de Boysson-Bardies, B., and Vihman, M. (**1991**). "Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants," Lang Speech **34**, 299–318.

Hallé, P., Segui, J., Frauenfelder, U., and Meunier, C. (**1998**). "Processing of illegal consonant clusters: A case of perceptual assimilation?," J. Exp. Psychol. Hum. Percept. Perform. **24**, 592–608.

Han, M. (**1962**). "Unvoicing of vowels in Japanese," Phonetic Studies **10**, 210–231.

Hughes, J. W. (**1946**). "The threshold of audition for short periods of stimulation," Proc. R. Soc. London, Ser. B **133**, 486–490.

Kahn, D. (**1980**). *Syllable-Based Generalizations in English Phonology* (Garland, New York).

Kawasaki, H. (**1982**). "An acoustical basis for universal constraints on sound sequences," Ph.D. dissertation, University of California, Berkeley.

Keating, P., and Huffman, M. (**1984**). "Vowel variation in Japanese," Phonetica **41**, 191–207.

Kenstowicz, M. (**2003**). "The role of perception in loanword phonology. A review of *Les emprunts linguistiques d'origine européenne en Fon*," Studies in African Linguistics **32**, 95–112.

Kewley-Port, D. (**1983**). "Time-varying features as correlates of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 606–608.

Kuhl, P., and Iverson, P. (**1995**). "Linguistic experience and the "perceptual magnet effect'," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange and J. Jenkins (York, Timonium, MD), pp. 121–154.

Laufer, A. (**1990**). "Hebrew (Illustrations of the IPA)," J. Int. Phonetic Assoc. **20**, 40–43.

Laufer, A. (**1998**). "Voicing in contemporary Hebrew in comparison with other languages," Hebrew Studies **39**, 143–179.

Lisker, L., and Abramson, A. (**1964**). "Cross-language study of voicing in initial stops: acoustical measurements," Word **20**, 384–422.

Mann, V., and Repp, B. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop consonant perception," Percept. Psychophys. **28**, 407–412.

Massaro, D., and Cohen, M. (**1983**). "Phonological context in speech perception," Percept. Psychophys. **34**, 338–348.

Moreton, E. (**2002**). "Structural constraints in the perception of English stop-sonorant clusters," Cognition **84**, 55–71.

Nearey, T., and Rochet, B. (**1994**). "Effects of place of articulation and vowel context on VOT production and perception for French and English stops," J. Int. Phonetic Assoc. **24**, 1–18.

New, B., Pallier, C., Ferrand, L., and Matos, R. (**2001**). "Une base de données lexicales du français contemporain sur internet: LEXIQUE^TM," ("An internet on-line lexical database for modern French: LEXIQUE^TM," Année Psychol. **101**, 447–462.

Obler, L. (**1982**). "The parsimonious bilingual," in *Exceptional Language and Linguistics*, edited by L. Obler and L. Menn (Academic, New York), pp. 339–346.

Padgett, J. (**1991**). "Structure in feature geometry," Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Paradis, C., and LaCharité, D. (**1997**). "Preservation and minimality in loanword adaptation," J. Linguistics **33**, 379–430.

Paradis, C., and LaCharité, D. (**2001**). "Guttural deletion in loanwords," Phonology **18**, 255–300.

Peterson, G., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Pitt, M. (**1998**). "Phonological processes and the perception of phonotactically illegal consonant clusters," Percept. Psychophys. **60**, 941–951.

Polivanov, E. (**1931**). "La perception des sons d'une langue étrangère," ("The perception of non-native language sounds"), Travaux du Cercle Linguistique de Prague **4**, 79–96.

Raphael, L., Tobin, Y., Faber, A., Most, T., Kollia, H., and Milstein, D. (**1995**). "Intermediate values of voice onset time," in *Producing Speech: Contemporary Issues for Katherine Harris*, edited by F. Bell-Berti and L. Raphael (AIP Press), pp. 117–127.

Raphael, L., Tobin, Y., and Most, T. (**1983**). "Atypical VOT categories in Hebrew and Spanish," J. Acoust. Soc. Am. **74**, S89A.

Rosen, H. (**1962**). *A Textbook of Israeli Hebrew* (The University of Chicago Press, Chicago, IL).

Saerens, M., Serniclaes, W., and Beeckmans, R. (**1989**). "Acoustic versus contextual factors in stop voicing perception in spontaneous French," Lang Speech **32**, 291–314.

Scharf, B. (**1978**). "Loudness," in *Handbook of Perception*, Hearing Vol. **4**, edited by E. Carterette and P. Morton (Academic, New York), pp. 187–242.

Selkirk, E. (**1988**). "Dependency, place, and the notion 'tier'," University of Massachusetts, Amherst (unpublished).

Shinohara, S. (**2006**). "Perceptual effects in final cluster reduction patterns," Lingua **116**, 1046–1078.

Silverman, D. (**1992**). "Multiple scansions in loanword phonology: Evidence from Cantonese," Phonology **9**, 289–328.

Simon, M. (**1967**). *Les Consonnes Françaises - Mouvements et Positions Articulatoires: La Lumière de la Cinématographie ("French consonants – articulatory movements and position: the contribution of cinematography"),* (Klincksieck, Paris).

Stevens, K., and Blumstein, S. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Téné, D. (**1972**). "The measured durations of Hebrew vowels - the schwa," in *Readings in Phonology*, edited by U. Ornan (Jerusalem Hebrew University Press, Jerusalem), pp. 239–251 (in Hebrew).

Tobin, Y. (**2002**). "Phonology as human behavior: Initial consonant clusters across languages," in *Signal, Meaning, and Message*, edited by W. Reid, R. Ortheguy, and N. Stern (Benjamin, Amsterdam), pp. 191–255.

Trubetzkoy, N. (**1939**). "Grundzüge der Phonologie," Travaux du Cercle Linguistique de Prague **7**, 1–271.

Vendelin, I., and Peperkamp, S. (**2004**). "Evidence for phonetic adaptation of loanwords: and experimental study," *Actes des Journées d'Etudes Linguistiques 2004*, pp. 129–131.

Yip, M. (**1993**). "Cantonese loanword phonology and optimality theory," J. East Asian Ling. **2**, 261–291.

Zawadzki, P. A. and Kuehn, D. P. (**1980**). "A cineradiographic study of static and dynamic aspects of American English /r/," Phonetica **37**, 253–266.

Zue, V. (**1976**). "Acoustic characteristics of stop consonants: A controlled study," Ph.D. dissertation. MIT, Cambridge, MA.

Effects of aging on word intelligibility and listening difficulty in various reverberant fields

Hayato Sato[a)]

Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501, Japan

Hiroshi Sato

Institute for Human Science & Biomedical Engineering, National Institute of Advanced Industrial Science and Technology, Tsukuba Central, Tsukuba, Ibaraki 305-8568, Japan

Masayuki Morimoto

Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501, Japan

Listening difficulty ratings is a new measure to evaluate speech transmission performance in public spaces. The advantage is that it can evaluate sound fields with higher speech transmission performance more accurately and sensitively than intelligibility scores. The effect of aging, or hearing level on listening difficulty ratings has not been clarified yet. In this paper, the results of the listening tests demonstrated that hearing levels above 2 kHz significantly affect both word intelligibility scores and listening difficulty ratings for the aged. For the aged, both of them can be estimated from the results for the young plus the average of hearing levels at 2 and 4 kHz. The equations for estimating the word intelligibility scores and the listening difficulty ratings for the aged are obtained from the multiple linear regression analyses. The estimated scores and ratings generally coincide with the measured scores and ratings, and the averaged difference between them is 5.2% for the word intelligibility scores and 3.8% for the listening difficulty ratings, in the range of listening tests performed in this paper. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2713715]

I. INTRODUCTION

Elderly people often find speech difficult to listen to, especially in reverberant and noisy conditions. The results of many studies indicate that the increase of hearing level is related to the difficulty. Hearing level of humans increases with aging, usually in the higher frequencies.[1] ISO 7029[2] documents the distributions of hearing level for varied age and sex in detail. Though Duquesnoy[3] reported that hearing level was not always a suitable measure for the hearing loss on intelligibility for an individual listener, the results averaged over a number of listeners showed that intelligibility scores correlated with hearing level at frequencies above 2 kHz.[4]

There are some studies on the effect of hearing level on intelligibility in reverberant fields. Duquesnoy and Plomp[5] measured speech reception thresholds (SRTs) for sentences in reverberant fields with the reverberation times from 0.4 to 2.3 s. The additional noise level was fixed at 52.5 dBA, and the speech level was varied to measure SRTs in units of the speech transmission index. The results showed that the SRT increased with increasing hearing level. Nábělek and Robinson[6] performed the modified rhyme test in reverberant fields. Six groups of listeners, whose mean ages were varied from 10 to 72 years and whose mean hear-

ing levels increased with increasing age, participated in the tests. The result of an analysis of variance showed that the main effect of age was statistically significant. Irwin and McAuley[7] obtained SRTs in units of the speech to noise ratio for two kinds of reverberant fields. SRTs for hearing-impaired listeners were significantly higher than those for normal listeners in both reverberant fields. The results for vowel and consonant recognition tests in reverberant fields also showed the high correlation between hearing level and the scores of the tests.[8,9] Estimating intelligibility scores for the aged from those for the young is useful for saving time and costs on assessments of speech transmission performance. The above-described knowledge indicates that the intelligibility scores for the aged can be estimated from their hearing levels.

Meanwhile, listening conditions in daily life are quite varied. It is important to investigate whether or not the effect of aging on speech perception varies in different sound fields. For example, Nábělek and Robinson[6] reported that an interaction between age and reverberation time was statistically significant: The effect of age increased with increasing reverberation time. However, intelligibility scores cannot discriminate among sound fields with higher speech transmission performance that are common in public spaces, because of its ceiling effect as a function of speech transmission performance.

a)Electronic mail: hayato@kobe-u.ac.jp

Morimoto *et al.*[10] developed the subjective measure of listening difficulty by using word lists with the highest word familiarity,[11] and demonstrated that it is more suitable than intelligibility scores for evaluating sound fields with higher speech transmission performance. Sato *et al.*[12] reported that word intelligibility scores are suitable for the speech to noise ratios less than 0 dB, while listening difficulty ratings are suitable for the speech to noise ratios from 0 to 15 dB, in case of young listeners. It is a constructive way to evaluate a wide range of listening conditions by using the two measures: Word intelligibility scores and listening difficulty ratings. However, previous tests of listening difficulty were not concerned with the effect of aging. It is not clear yet how hearing level and aging affect listening difficulty ratings.

The present study has two purposes: (1) To clarify the effect of aging on word intelligibility and listening difficulty in various reverberant fields, with a focus on hearing level; and (2) to provide the method for estimating word intelligibility scores and listening difficulty ratings for the aged from those for the young and hearing levels.

II. METHOD

In the present study, test words, test sound fields, and procedures were intended to reproduce those used in Morimoto's work[10] as much as possible.

A. Test word and sound field

The word lists developed by Sakamoto *et al.*[11] were used in the listening test. Each word list consists of 50 Japanese words, of similar word familiarity and including phonetic balance. Each word has four syllables. Word familiarity is a subjective variable, which indicates how familiar native speakers are with the word. It had been rated from 1.0 (minimum) to 7.0 (maximum) by 32 young adults. The word lists are divided into four word familiarity levels. One word list with the highest level, corresponding to the word familiarity from 7.0 to 5.5, was used. Test words were spoken by a female talker, and recorded in an anechoic room.

The reverberant sounds used in the present study were set to be detrimental for speech perception, in order to discuss simply the relation between the effect of hearing level on speech perception and reverberant sounds. The test words were preliminarily convolved with impulse responses using software on a computer to add reverberant sounds. Figure 1 shows schematically the impulse response used in the listening test. The impulse response was composed of a direct sound followed by a reverberant decay. The reverberant decay always started at 50 ms after the direct sound to make all reverberant sounds detrimental for speech perception. Early reflections were not added to avoid the beneficial effect of them on speech perception, which was reported by Bradley *et al.*[13] The parameters were reverberation time and sound pressure ratio of the onset component of the reverberant signal to the direct sound described as Pr/Pd. The reverberation time was set at 0.5, 0.9, 2.0, and 6.0 s. The frequency characteristic of the reverberation time was flat. The Pr/Pd was set at 0.5, 0.25, and 0.1. A total of eight sound fields were used in the listening test (see Table III).
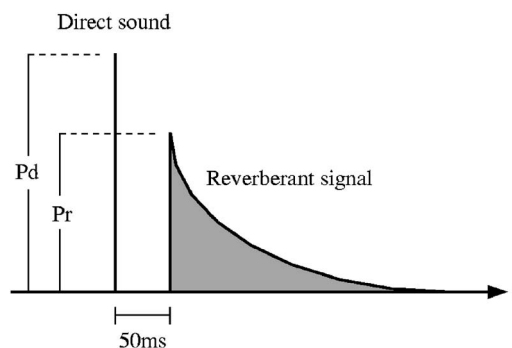
FIG. 1. Impulse response of the test sound field. Pd and Pr are the sound pressure level of a direct sound and the onset of a reverberation signal, respectively.

B. Listener

A total of 142 listeners participated in the listening tests. They varied in age from 56 to 79 years old. All of them could live independently and actively in societies, and were suitable as samples of typical elderly people. Hearing levels for both ears were measured using an audiometer in a small soundproof room. Figure 2 represents the mean hearing level and ±1 s.d. between the listeners at each frequency. The mean hearing level increased with increasing frequency. The variance among the listeners also increased with increasing frequency. ISO 7029[2] enables us to calculate the mean hearing level using the listeners' age and sex. The mean hearing level for the listeners in the present study was estimated from ISO 7029 to confirm that hearing levels of the listeners were not far from the standardized data. The double-dashed line in Fig. 2 represents the estimated mean hearing levels from ISO 7029. The large differences are shown at frequencies less than 500 Hz, and the differences increased with decreasing frequency. Sound insulation performance of the soundproof room was not good enough for audiometric testing especially at lower frequencies less than 500 Hz. Therefore, the differences might be due to background noise in the room where the soundproof room was installed. Meanwhile, the mean hearing levels at high frequencies are close to the estimated
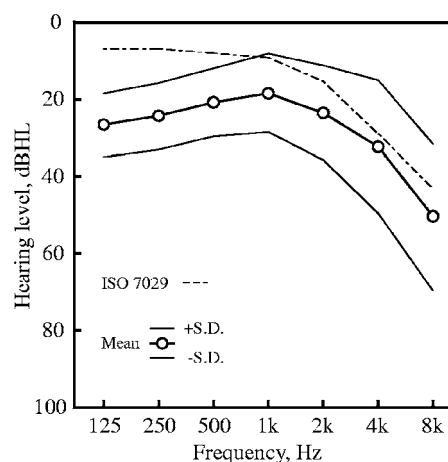
FIG. 2. Audiogram for listeners used in the listening test. Open circles and two solid lines represent the mean and ±1 s.d. of the hearing level of listeners, respectively. Dashed line represents the estimated mean hearing level of listeners by ISO 7029.

TABLE I. Categories of listening difficulty.

1	Not	difficult
2	A little	difficult
3	Fairly	difficult
4	Extremely	difficult

mean hearing levels. Therefore, hearing levels of the listeners in the present study are generally typical for their age.

C. Procedure

A total of 400 test signals (50 words × 8 sound fields) were prepared and were divided into ten sets of 40 test signals. Each set included all sound fields evenly: 5 test signals per sound field, and did not include the same test word. Each listener listened to a set of test signals in a random order. Therefore, a total of 710 responses (5 test signals × 142 listeners) were obtained for each sound field in the whole of the listening test. Though the test words used in the present study had the phonetic balance, it was difficult to keep the phonetic balance for each listener and each sound field in this procedure.

The test signals were presented to two listeners at a time from a loudspeaker in an anechoic room. The listening positions were located at 2 m distances from the loudspeaker and at ±30° from the central axis of the loudspeaker. Listeners were forced to face the loudspeaker. The test signals were presented from a single loudspeaker to avoid the beneficial effect of binaural listening in reverberant sound fields, and to discuss simply the result of the listening test. The sound pressure levels of the test signals were set at 65 dBA, fast, peak on the average of all test words at listening positions when only the direct sound was presented. Each listener was asked to repeat by writing down each word in Kana orthography as they listened and simultaneously to rate the listening difficulty into one of the four categories shown in Table I.

III. RESULTS AND DISCUSSIONS

Listeners were divided into several groups according to their hearing levels in 10 dBHL steps, which were averaged over both ears. The classifications were made for each hearing level at frequencies from 250 Hz to 4 kHz, and also for the average of hearing levels at 2 and 4 kHz. Table II shows the number of listeners for each group. Word intelligibility scores and listening difficulty ratings were obtained from the results of the listening test. The word intelligibility score is the percentage of the test words answered correctly. The listening difficulty rating is the percentage of responses that indicated some level of difficulty, that is, responses from "2" to "4" in Table I. Note that listening difficulty ratings decrease when conditions for speech improved, contrary to word intelligibility scores. Word intelligibility scores, listening difficulty ratings, and mean hearing levels were obtained for each group and each sound field. However, groups that consisted of only one listener were not included in the following analyses.

A. Word intelligibility scores

Figure 3 represents the word intelligibility scores as a function of hearing level. Each panel represents those for each frequency at which hearing level was measured, except the lower right panel represents the word intelligibility scores as a function of the average of hearing levels at 2 and 4 kHz. Different symbols represent different sound fields. Roughly speaking, the word intelligibility scores decreased with increasing the reverberation time and increasing Pr/Pd. This behavior coincides with that for the young obtained by Morimoto et al.[10] Meanwhile, the word intelligibility scores decreased with increasing hearing level. This behavior is apparent for 1 kHz and higher frequencies, and also for the average of 2 and 4 kHz.

A repeated measures two-way analysis of variance with "hearing level" as a between subject factor and "sound field" as a within subject factor was performed for each frequency from 250 Hz to 4 kHz and for the average of 2 and 4 kHz. The significant level was set at 5%. The results of the analyses revealed that (1) the main effect of hearing level was significant for all frequencies, (2) the main effect of sound field was significant for all frequencies, and (3) the interaction between hearing level and sound field was significant only for 250 Hz.

Table III represents F-ratios of simple main effects of hearing level for each sound field and each frequency.

TABLE II. The number of listeners for the groups that are classified by hearing level in 10 dBHL steps.

Hearing level[a] (dBHL)	Frequency (Hz)					
	250	500	1000	2000	4000	2000–4000
0–9	2	5	18	10	7	5
10–19	32	61	69	51	30	44
20–29	69	47	38	44	33	37
30–39	30	23	9	16	23	26
40–49	8	5	5	13	21	17
50–59	1	1	3	7	15	11
60–69	0	0	0	1	12	1
70–79	0	0	0	0	0	1
80–89	0	0	0	0	1	0

[a]Average of the left and the right ears.

FIG. 3. Word intelligibility scores as a function of hearing level and a parameter of the sound fields, for each frequency at which the hearing level was measured. The word intelligibility scores and hearing level are the average for the group classified according to hearing level in 10 dBHL steps. Different symbols represent different sound fields.
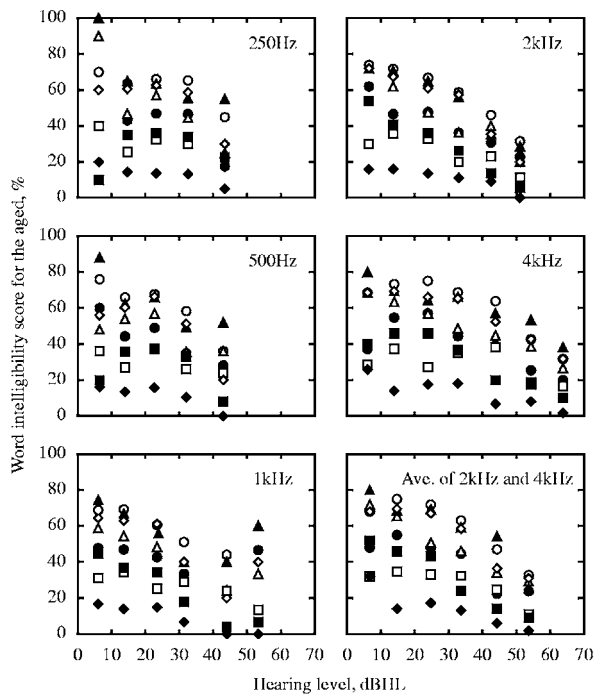
○: (a) RT=0.5s, Pr/Pd=0.50
●: (b) RT=0.9s, Pr/Pd=0.50   △: (c) RT=0.9s, Pr/Pd=0.25   ▲: (d) RT=0.9s, Pr/Pd=0.10
□: (e) RT=2.0s, Pr/Pd=0.50   ■: (f) RT=2.0s, Pr/Pd=0.25   ◇: (g) RT=2.0s, Pr/Pd=0.10
◆: (h) RT=6.0s, Pr/Pd=0.50

FIG. 4. As for Fig. 3, but for listening difficulty ratings.

Simple main effects for the sound field (h) were not significant for all frequencies except for the average of 2 and 4 kHz, because reverberant sounds for the sound field (h) decreased the word intelligibility scores to less than 20% even for the aged with lowest hearing level. The F-ratio tends to increase with increasing frequency, regardless of sound fields. Especially the F-ratios for 2 and 4 kHz were higher than those for lower frequencies. This means that word intelligibility scores in reverberant fields are mainly affected by hearing levels at 2 and 4 kHz. Although the F-ratio tends to increase with decreasing reverberation time

and decreasing Pr/Pd regardless of frequency, the interactions between hearing level and sound field were not significant for almost all frequencies. This means that the effect of hearing level on word intelligibility scores seems to be constant, regardless of sound fields.

B. Listening difficulty ratings

Figure 4 represents the listening difficulty ratings as a function of hearing level. For 250 and 500 Hz, hearing level did not affect the listening difficulty ratings, while the listening difficulty ratings increased (i.e., conditions for speech deteriorated) with increasing the reverberation time and Pr/Pd. For 1 kHz and higher frequencies and the average of 2 and 4 kHz, the effect of hearing level seems to increase

TABLE III. F-ratios of the simple main effect of "hearing level" on word intelligibility scores for each sound field. The values were obtained by a repeated measures two-way analysis of variance with "hearing level" and "sound field." The analyses of variance were performed for the hearing level at each frequency.

Sound field	RT (s)	Pr/Pd	Frequency (Hz)					
			250	500	1000	2000	4000	2000–4000
a	0.5	0.50	1.12	4.01*	1.67	7.19*	8.63*	7.40*
b	0.9	0.50	1.76	2.67*	1.25	5.12*	5.91*	5.23*
c	0.9	0.25	6.66*	1.72	2.38*	9.13*	6.57*	7.81*
d	0.9	0.10	4.06*	4.04*	2.68*	9.01*	5.44*	9.78*
e	2.0	0.50	0.53	0.63	0.72	2.19	2.45*	2.22
f	2.0	0.25	1.47	2.73*	3.86*	8.35*	6.59*	8.92*
g	2.0	0.10	2.20	5.67*	4.27*	10.75*	6.72*	8.00*
h	6.0	0.50	0.34	0.76	0.77	0.95	2.09	3.01*

*Significant at 5% levels.

TABLE IV. As for Table III, but for listening difficulty ratings.

Sound field	RT (s)	Pr/Pd	Frequency (Hz)					
			250	500	1000	2000	4000	2000–4000
a	0.5	0.50	1.07	0.15	2.80[*]	7.13[*]	6.20[*]	6.26[*]
b	0.9	0.50	0.12	0.26	0.98	0.98	1.33	1.08
c	0.9	0.25	0.08	0.89	1.61	3.23[*]	2.55[*]	2.02
d	0.9	0.10	7.00[*]	1.79	4.19[*]	6.41[*]	4.94[*]	5.83[*]
e	2.0	0.50	0.40	0.95	0.33	0.11	0.30	0.18
f	2.0	0.25	0.71	0.32	0.92	0.89	0.75	0.67
g	2.0	0.10	3.33[*]	0.80	2.09	5.40[*]	3.94[*]	3.12[*]
h	6.0	0.50	0.02	0.04	0.04	0.03	0.14	0.11

[*]Significant at 5% levels.

with decreasing reverberant sound energy. The listening difficulty ratings for the sound fields with less reverberant sound energy, for example, the sound field (a) which is represented by open circles in Fig. 4, increased with increasing hearing level. However, the listening difficulty ratings for the sound fields with longer reverberation time and higher Pr/Pd were constant at a high value. Consequently, the effect of the sound fields decreased with increasing hearing level.

A repeated measures two-way analysis of variance was performed as for the word intelligibility scores. The results revealed that (1) the main effect of hearing level was significant for 2 and 4 kHz and the average of 2 and 4 kHz, (2) the main effect of sound field was significant for all frequencies, and (3) the interaction between hearing level and sound field was significant for 2 and 4 kHz and the average of 2 and 4 kHz. Table IV represents F-ratios of simple main effects of hearing level for each sound field and each frequency. As for the case of word intelligibility scores, significant simple main effects were found for 2 and 4 kHz. However, simple main effects were significant only for the four sound fields of (a), (c), (d), and (g) where the energy of reverberant sounds was less. This result means that the effect of hearing level on listening difficulty ratings is smaller than that on word intelligibility scores. One possible reason for this result is the ceiling effect of listening difficulty ratings as a function of hearing level. Listening difficulty ratings for all listeners ranged between 78% and 97% for the sound fields where the simple main effects were not significant. Therefore, the listening difficulty ratings did not change so much according to hearing level. The F-ratio tends to increase with decreasing the reverberation time and decreasing Pr/Pd regardless of frequency. This tendency corresponds to the significant interactions between hearing level and sound field for 2 and 4 kHz. This result coincides with the result of Morimoto et al.[10] that word intelligibility scores can evaluate the speech transmission performance more sensitively than listening difficulty ratings for sound fields with lower speech transmission performance. The speech transmission performances for half of the test sound fields in the present study were too low even for the aged with the lowest hearing level to evaluate using listening difficulty ratings.

C. Comparison between word intelligibility scores for the young and those for the aged

Morimoto et al.[10] obtained word intelligibility scores and listening difficulty ratings for the young by using the same sound fields as those used in the present study. Hearing levels for the young listeners were not measured, but they did not report any known hearing problems. Figure 5 represents the relations between the word intelligibility scores for the young by Morimoto et al. and those for the aged in the present study for each group that was classified by the average of hearing levels at 2 and 4 kHz. There are two reasons for adopting the averaged hearing level: (1) It significantly affects both subjective measures in reverberant fields, and (2) it varies widely for the aged. Because they are expected to be normally distributed, the word intelligibility scores were transformed to z values; the word intelligibility score of 50% corresponds to the z value of zero. Word intelligibility scores less than 5% were omitted to prevent an excessive effect of it on a linear regression analysis.
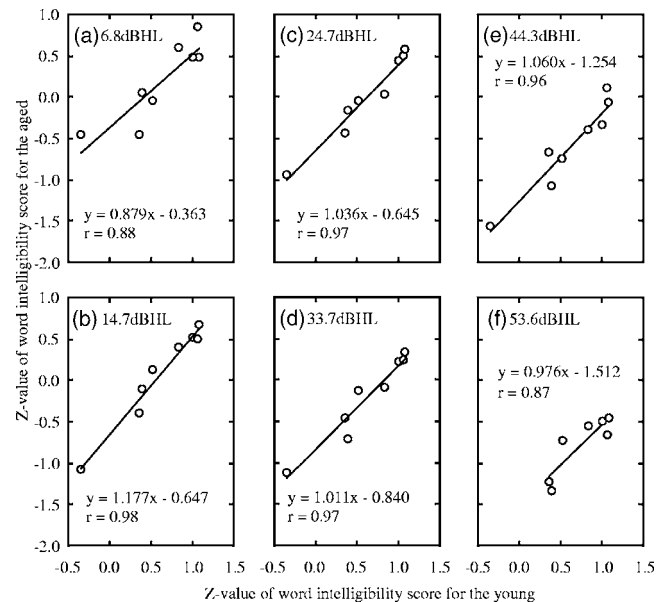
FIG. 5. Z values of word intelligibility scores for the aged as a function of those for the young for each group classified by the average of hearing levels at 2 and 4 kHz. The word intelligibility scores for the young are obtained by Morimoto et al.
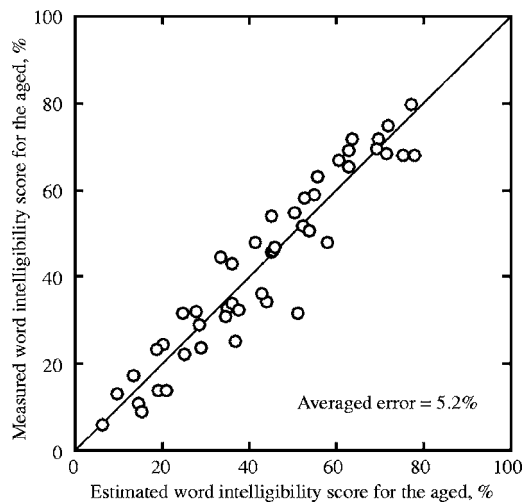
FIG. 6. The relation between estimated and measured word intelligibility scores for the aged.

FIG. 7. Estimated word intelligibility scores for the aged as a function of word intelligibility scores for the young and as a parameter of hearing level.

The z value of word intelligibility score for the aged strongly correlates with that for the young. The correlation coefficients between them are more than 0.87 for all hearing levels. This means that the word intelligibility scores for the aged can be estimated from those for the young. Note that the regression coefficients are almost constant regardless of hearing level. This behavior seems to correspond to the results of the analysis of variance that the interaction between hearing level and sound field was not significant, given that the difference of sound fields is replaced by the difference of word intelligibility scores for the young. Meanwhile, the constant terms of the regression equations inversely correlate with hearing level. This means that the word intelligibility scores for the aged can also be estimated from hearing level.

A multiple linear regression analysis was performed to estimate z values of word intelligibility scores for the aged from the two independent variables: WI_Y, the word intelligibility score for the young, and HL, the average of hearing levels at 2 and 4 kHz. Based on the analysis, $z(WI_A, HL)$, the z value of the word intelligibility score for the aged with hearing level of HL, can be estimated as follows:

$$z(WI_A, HL) = 1.02z(WI_Y) - 0.0232HL - 0.184, \qquad (1)$$

where $z(x)$ is the z value of x. The coefficient of determination is 0.91, and this demonstrates that Eq. (1) can estimate word intelligibility for the aged with high accuracy. The partial regression coefficients for both variables and the constant term are statistically significant ($p < 0.05$). The standardized partial regression coefficients are 0.76 and −0.63 for $z(WI_Y)$ and HL, respectively. This means that hearing level affects the word intelligibility scores for the aged as much as those for the young, or the sound fields. The sum of the HL term and the constant term represent the difference between the z value of the word intelligibility scores for the aged and those for the young, because the partial regression coefficient for $z(WI_Y)$ is approximately 1.0. That the constant term is statistically significant indicates that some factors other than hearing level decrease the z value of the word intelligibility score for the aged. In this regard, the decrease of the word intelligibility score due to
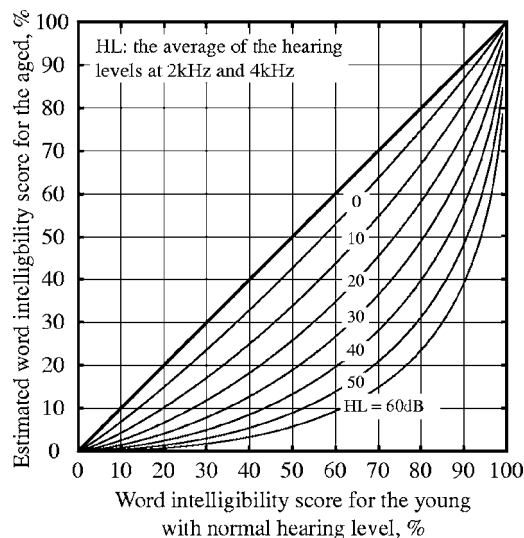
the constant term is not so much. The difference of z value of 0.184 corresponds to the difference of the word intelligibility score of only 7% at a maximum.

Figure 6 represents the relation between the measured scores for the aged and the estimated ones for the aged by Eq. (1) and an inverse z-transformation. The estimated scores generally coincide with the measured scores, and the averaged difference between them is 5.2%. Figure 7 represents the relation between the word intelligibility scores for the young and those for the aged with varied hearing levels based on Eq. (1). This chart is useful to estimate how much the decrease of word intelligibility scores for the aged will be.

D. Comparison between listening difficulty ratings for the young and those for the aged

Figure 8 represents the z value of listening difficulty ratings for the young by Morimoto et al.[10] and those for the aged in the present study. As for the word intelligibility scores, the listening difficulty ratings for the aged were obtained for each group classified by the average of hearing levels at 2 and 4 kHz in 10 dBHL steps, and were transformed to z values. The listening difficulty ratings more than 95% were omitted to prevent an excessive effect of it on a linear regression analysis. The z values of the listening difficulty ratings for the aged strongly correlate with those for the young, except the correlation coefficient for the highest hearing level is relatively lower than those for lower hearing levels [see Fig. 8(f)].

Both the regression coefficients and the constant terms of the regression equations correlate with hearing level. The regression coefficients are inversely proportional to hearing level ($r = -0.92$), and the constant terms are directly proportional to the hearing level ($r = 0.98$). The latter result is the same behavior as that for word intelligibility scores. However, the former result is different from that for word intelligibility scores. A possible reason for the former result is the ceiling effect of the listening difficulty ratings as a function
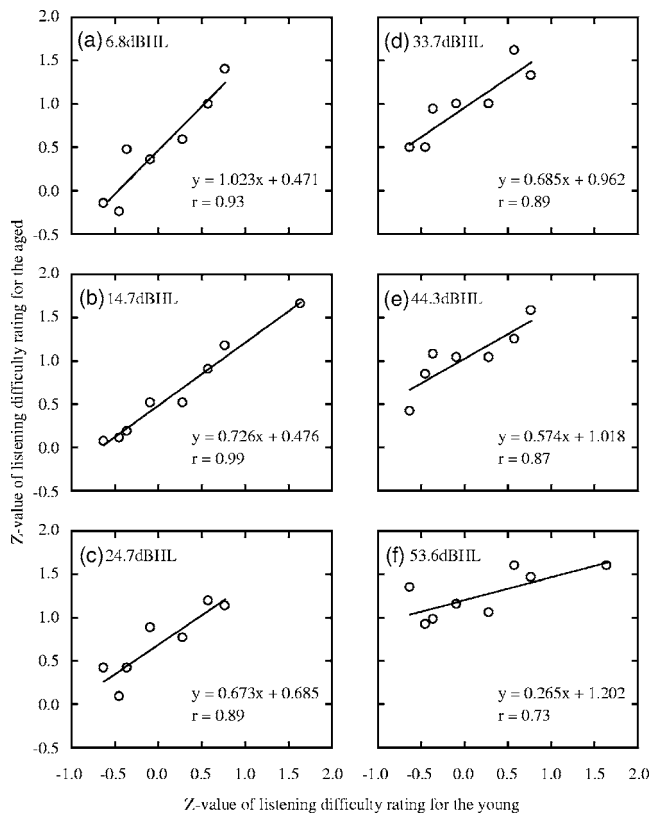
FIG. 8. As for Fig. 5, but for listening difficulty ratings.

FIG. 9. As for Fig. 6, but for listening difficulty ratings.

of hearing level. Hearing level did not affect the listening difficulty ratings when reverberant sounds dramatically increased the ratings.

These results indicate that the listening difficulty ratings for the aged can be estimated from the listening difficulty ratings for the young and hearing levels. If a linear regression analysis is adopted, $z(LD_A, HL)$, the z value of the listening difficulty rating for the aged with hearing level of HL, can be estimated as follows:

$$z(LD_A, HL) = (a_1 HL + a_2)z(LD_Y) + b_1 HL + b_2, \quad (2)$$

where $z(x)$ is z value of x and LD_Y is listening difficulty rating for the young. However, Eq. (2) can be transformed as follows:

$$z(LD_A, HL) = A_0 + A_1 z(LD_Y) + A_2 HL + A_3 z(LD_Y)HL. \quad (3)$$

Therefore, a multiple linear regression analysis was performed to determine Ax: the regression coefficients and the constant term in Eq. (3). Equation (4) was obtained from the analysis,

$$z(LD_A, HL) = 0.298 + 0.979z(LD_Y) + 0.0168HL$$
$$- 0.0123z(LD_Y)HL. \quad (4)$$

The coefficient of determination is 0.86. The standardized partial regression coefficients are 1.18, 0.56, and −0.53 for $z(LD_Y)$, HL, and $z(LD_Y)$HL, respectively. The partial regression coefficients and the constant term are statistically significant ($p < 0.05$). Although the correlation coefficient of 0.87 between $z(LD_Y)$ and $z(LD_Y)$HL is too high to in-

clude both variables in the analysis, the $z(LD_Y)$HL term should be included to adapt the behavior of the estimated ratings to that of the measured ratings that the effect of $z(LD_Y)$ decreases with increasing hearing level. Equation (4) indicates that the z value of the listening difficulty rating for the aged increases around 0.3 over that for the young, when hearing level is 0 dBHL. This means that there are some factors other than hearing level that affect the z value of the listening difficulty ratings for the aged. Note that the constant term for listening difficulty ratings is higher than for word intelligibility scores. This result may be explained by the difference of sensitivity between the two measures. As demonstrated in the present study and a previous study,[10] listening difficulty ratings are more sensitive than word intelligibility scores when the speech transmission performance is relatively high. Therefore, the listening difficulty ratings are affected more strongly by factors other than hearing level.

Figure 9 represents the relation between the measured ratings for the aged and the estimated ones for the aged by Eq. (4) and an inverse z-transformation. The estimated ratings generally coincide with the measured ratings, and the
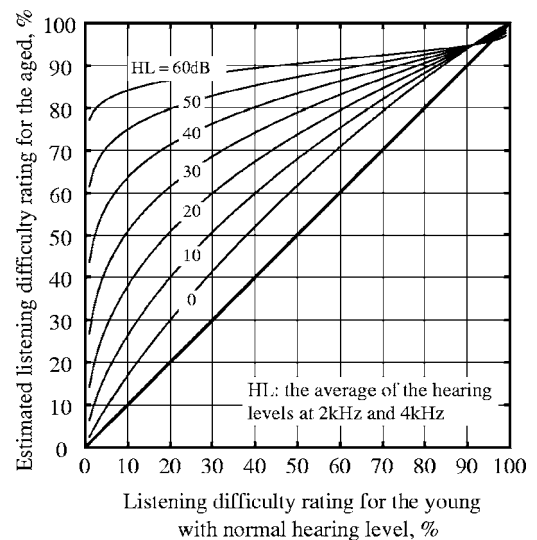
FIG. 10. As for Fig. 7, but for listening difficulty ratings.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Sato *et al.*: Effects of aging on listening difficulty   2921

averaged difference between them is 3.8%. Figure 10 represents the relation between the listening difficulty ratings for the young and those for the aged with various hearing levels based on Eq. (4). This chart clearly represents the increase of listening difficulty ratings due to aging, except the estimated ratings for the aged crossed when they exceed around 95%. This unexpected behavior is caused by the $z(LD_y)HL$ term, but the estimation of such high listening difficulty ratings is not important from a practical point of view.

IV. CONCLUSIONS

The effect of hearing levels on word intelligibility scores and listening difficulty ratings for the aged in various reverberant fields were investigated. The results of the listening test clarified that (1) hearing levels above 2 kHz significantly affect both word intelligibility scores and listening difficulty ratings for the aged; and (2) the effect of hearing level on listening difficulty ratings is smaller than that on word intelligibility scores because of the ceiling effect of listening difficulty ratings as a function of hearing level. The word intelligibility scores and the listening difficulty ratings for the aged were compared to those for the young that were obtained by Morimoto *et al.*

The results of the comparison clarified that (1) the word intelligibility scores for the aged strongly correlate with those for the young, regardless of hearing level for the aged, and the same holds for the listening difficulty ratings and (2) both word intelligibility scores and listening difficulty ratings for the aged can be estimated from those for the young and hearing level with high accuracy.

[1]L. J. Brant and J. L. Fozard, "Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging," J. Acoust. Soc. Am. **88**, 813–820 (1990).

[2]ISO 7029:2000, *Acoustics - Statistical Distribution of Hearing Thresholds as a Function of Age* (International Organization for Standardization, Geneva, 2000).

[3]A. J. Duquesnoy, "The intelligibility of sentences in quiet and in noise in aged listeners," J. Acoust. Soc. Am. **74**, 1136–1144 (1983).

[4]K. D. Kryter, C. Williams, and D. M. Green, "Auditory acuity and the perception of speech," J. Acoust. Soc. Am. **34**, 1217–1223 (1962).

[5]A. J. Duquesnoy and R. Plomp, "Effect of reverberation and noise on the in cases of prebyscusis," J. Acoust. Soc. Am. **68**, 537–544 (1980).

[6]A. K. Nábělek and P. K. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248 (1982).

[7]R. J. Irwin and S. F. McAuley, "Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation," J. Acoust. Soc. Am. **81**, 1557–1565 (1987).

[8]A. K. Nábělek, "Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss," J. Acoust. Soc. Am. **84**, 476–484 (1988).

[9]K. S. Helfer and R. A. Huntley, "Aging and consonant errors in reverberation and noise," J. Acoust. Soc. Am. **90**, 1786–1796 (1991).

[10]M. Morimoto, H. Sato, and M. Kobayashi, "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," J. Acoust. Soc. Am. **116**, 1607–1613 (2004).

[11]S. Sakamoto, Y. Suzuki, S. Amano, K. Ozawa, T. Kondo, and T. Sone, "New lists for word intelligibility test based on word familiarity and phonetic balance," J. Acoust. Soc. Jpn. **54**, 842–849 (1998) (in Japanese).

[12]H. Sato, J. Bradley, and M. Morimoto, "Using listening difficulty ratings of conditions for speech communication in rooms," J. Acoust. Soc. Am. **117**, 1157–1167 (2005).

[13]J. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**, 3233–3244 (2003).

2922   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Sato *et al.*: Effects of aging on listening difficulty

Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech

Paula C. Stacey[a)]

Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom
and MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

A. Quentin Summerfield

Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom

Five experiments were designed to evaluate the effectiveness of "high-variability" lexical training in improving the ability of normal-hearing subjects to perceive noise-vocoded speech that had been spectrally shifted to simulate tonotopic misalignment. Two approaches to training were implemented. One training approach required subjects to recognize isolated words, while the other training approach required subjects to recognize words in sentences. Both approaches to training improved the ability to identify words in sentences. Improvements following a single session (lasting 1–2 h) of auditory training ranged between 7 and 12 %pts and were significantly larger than improvements following a visual control task that was matched with the auditory training task in terms of the response demands. An additional three sessions of word- and sentence-based training led to further improvements, with the average overall improvement ranging from 13 to 18 %pts. When a tonotopic misalignment of 3 mm rather than 6 mm was simulated, training with several talkers led to greater generalization to new talkers than training with a single talker. The results confirm that computer-based lexical training can help overcome the effects of spectral distortions in speech, and they suggest that training materials are most effective when several talkers are included. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2713668]

I. INTRODUCTION

Cochlear implantation improves the speech perception abilities of postlingually deafened adults with profound-to-total hearing loss (e.g., Summerfield and Marshall, 1995). However, outcomes following implantation are highly variable (Gantz *et al.*, 1993). One way to improve the speech-perception skills of adult cochlear-implant users might be to administer auditory training. Initial investigations (Gagne *et al.*, 1991; Dawson and Clark, 1997; Busby *et al.*, 1991) provided only limited evidence in support of the effectiveness of such training. However, more recent studies have shown that computer-based auditory training can improve both timbre-recognition (Gfeller *et al.*, 2002) and speech-perception skills (Fu *et al.*, 2005a) of adults who use implants. Fu *et al.* (2005a) argued that previous studies had failed to find systematic benefits because insufficient training had been provided. Extensive training can now be delivered at low cost via personal computers. As a precursor to working with patients, the experiments reported here evaluated the effectiveness of two computer-based training approaches with normal-hearing subjects who listened to speech through a simulation of the information provided by a cochlear-implant system (Shannon *et al.*, 1995).

A variable that has been associated with poor speech perception by users of cochlear implants is tonotopic mis-alignment between the frequency band transmitted by an electrode and the characteristic frequency of the location of that electrode (Skinner *et al.*, 2002; Yukawa *et al.*, 2004). One reason for providing auditory training would be to help overcome the difficulties in speech perception caused by tonotopic misalignment. The consequences of tonotopic misalignment for speech perception can be investigated with normal-hearing subjects who listen to speech through noise-band vocoders designed to simulate the information provided by a cochlear implant (Shannon *et al.*, 1995; Baskent and Shannon, 2003; Dorman *et al.*, 1997; Rosen *et al.*, 1999). Accuracy of speech recognition declines when signals are spectrally shifted to simulate tonotopic misalignment (Baskent and Shannon, 2003; Dorman *et al.*, 1997; Rosen *et al.*, 1999). However, the decline can be ameliorated with auditory training (Rosen *et al.*, 1999; Fu *et al.*, 2005b; Faulkner *et al.*, 2006).

Rosen *et al.* (1999) tested the ability of normal-hearing subjects to perceive spectrally shifted speech that had been processed to simulate the consequences of tonotopic misalignment of 6.5 mm. Fewer than 1% of words in sentences were identified correctly, compared with 64% correct performance with unshifted signals. However, after nine 20-min sessions of auditory training using connected discourse tracking (CDT, De Filippo and Scott, 1978), performance improved to 30% correct. In CDT, an experimenter reads a passage of text, and the subject attempts to repeat verbatim what was said, with corrective feedback from the reader. Although CDT is an effective training approach (see also

[a)]Current address: Division of Psychology, Nottingham Trent University, Burton Street, Nottingham, NG1 4BU, United Kingdom. Electronic mail: paula.stacey@ntu.ac.uk

Faulkner *et al.*, 2006), it is labor intensive and expensive to administer clinically. Self-administered techniques could be a cost-effective alternative. Fu *et al.* (2005b) compared the effectiveness of two self-administered computer-based training protocols in improving the ability of normal-hearing listeners to discriminate spectrally-distorted vowel and consonant sounds. "Word-based training" required subjects to identify the vowels in consonant-vowel-consonant monosyllabic words; "sentence-based training" consisted of a computer-based CDT procedure. Both training approaches led to significant improvements in the ability to discriminate consonant sounds, but while word-based training led to significant improvements in the ability to discriminate vowel sounds, sentence-based training did not. Fu *et al.* (2005b) therefore concluded that word-based training might be more effective than sentence-based training in developing the speech-perception skills of cochlear-implant users.

There is a limitation with the study by Fu *et al.* (2005b) however, since they did not include a test of sentence perception. It is possible that sentence-based training leads to larger improvements in sentence perception than does word-based training. In addition, performance on a test of sentence perception might provide a more representative test of subjects' ability to communicate in everyday situations. Similar to Fu *et al.* (2005b), the present experiments examined the effectiveness of two computer-based lexical training approaches. The first training task required subjects to recognize isolated words, while the second training task required subjects to recognize words in sentences. The effectiveness of auditory training was assessed using tests of consonant and vowel discrimination. In addition, we assessed the extent to which each of the training tasks led to improvements on a test of sentence perception, thereby addressing the limitation of the study by Fu *et al.* (2005b). The experiments addressed two further issues, discussed below.

A. High-variability training

Studies of perceptual learning for speech have shown that training with several talkers, usually termed 'high-variability' training, is more effective than training with a single talker. High-variability lexical training is effective in training Japanese Americans to distinguish /r/ and /l/ (Lively *et al.*, 1993, 1994; Logan *et al.*, 1991; Bradlow *et al.*, 1997), for native English speakers to perceive Cantonese-accented speech (Bradlow and Bent, 2003), for American English speakers learning to classify American dialects (Clopper and Pisoni, 2004), and for North American subjects seeking to learn Mandarin tones (Wang *et al.*, 1999). These studies have demonstrated that training with several talkers is more generalizable to new talkers compared with training with a single talker. Training with many talkers may help subjects to dissociate talker-specific information from lexically specific information, since the two sorts of information are confounded when only one talker is used in training. The present experiments investigated whether high-variability training is also advantageous in training subjects to perceive spectrally distorted speech.

B. Training-related versus incidental learning

Although improvements on tests of speech perception following auditory training may be caused by the training task (training-related learning), the contribution of "incidental learning" must also be considered. Incidental learning refers to improvements that occur independent of the auditory training task, through procedural learning of task demands (Robinson and Summerfield, 1996), or perceptual learning resulting from repeating exposure to test materials. A further type of incidental learning has also been documented. Amitay *et al.* (2006) reported larger improvements in frequency discrimination for control subjects who played a purely visual computer game between successive tests than for control subjects who did not engage in an intervening task. These results suggest that maintaining attention and arousal, without explicit training, may be sufficient to lead to improvements on perceptual tasks. In order to evaluate the extent to which a training task has contributed to improvements in performance, it is important to factor out improvements related to "incidental learning."

Rosen *et al.* (1999) did not include a control condition, so it was not possible to measure the extent to which auditory training, rather than incidental learning, led to the observed improvements in performance. Fu *et al.* (2005b) did include a control condition, in which one group of subjects undertook repeated tests of speech perception without undertaking training. However, this procedure does not control the effects of maintaining attention and arousal. In our experiments, we controlled incidental learning by comparing improvements following auditory training with improvements following a "matched" visual control task that exposed subjects to the same vocabulary and imposed similar task demands as the auditory training task. We use the term "matched" to refer to our goal of matching the visual control task with the auditory training task in terms of its content, nature, and task demands.

C. Aims and key hypotheses

Against this background, five experiments are reported in this paper. Experiments 1–3 examined whether a single session (lasting 1 to 2 h) of auditory training improved the perception of spectrally shifted noise-vocoded speech more than did a matched visual control task. These experiments also tested the hypothesis that high-variability training is more effective than training provided by a single talker. A word-based training task was used in experiments 1 and 2, while a sentence-based training task was used in experiment 3. Experiments 1 and 2 differed in the degree to which signals were spectrally shifted. In experiment 1, speech was spectrally shifted to simulate a 6-mm tonotopic misalignment, while experiment 2 simulated a 3-mm tonotopic misalignment. The aim was to measure the relationship between the degree of tonotopic misalignment and the amount of improvement. Experiments 4 and 5 included four high-variability training sessions to establish whether performance continues to improve throughout 4 to 6 h of word- (experiment 4) and sentence-based (experiment 5) training.

In summary, the experiments addressed three hypotheses:

(1) Word-based (experiments 1 and 2) and sentence-based auditory training (experiment 3) lead to larger improvements on tests of sentence, consonant, and vowel perception than does a matched visual control task.
(2) High-variability training leads to larger improvements in speech perception than does single-talker training (experiments 1–3).
(3) Extending the amount of word-based (experiment 4) and sentence-based (experiment 5) auditory training leads to continued improvements in performance.

II. METHODS COMMON TO ALL EXPERIMENTS

A. Design and procedure

The overall design and procedure were the same in each experiment. Subjects initially completed baseline tests of speech perception, followed by a variable number of sessions of auditory training, followed by further tests of speech perception. The two training procedures are described in the methods sections of experiments 1 and 3. In all experiments, subjects completed an IEEE sentence test, then a test of vowel discrimination, followed by a test of consonant discrimination. These tests are described below. Training and testing took place in a double-walled sound-attenuated chamber. No feedback on accuracy was provided during testing.

Speech materials used for training and for tests of speech perception were recorded digitally (sample rate 44.1 kHz, amplitude quantification 16 bits) in a carpeted double-walled sound attenuated chamber. The talkers had a range of British and Irish accents. Male talkers were aged between 27 and 58 years (mean age 44 years); female talkers were aged between 27 and 40 years (mean age 30 years); and child talkers were aged 7 and 8 years. Stimuli were presented through an Audiomaster LS3/5A loudspeaker. Peak stimulus levels, measured with a Bruel and Kjaer Type 2260 sound level meter using a 1-s integration time and a Type 4189 half-inch microphone, ranged between 65 and 75 dB(A) across training and test materials.

B. IEEE sentence test

Four blocks of 80 sentences from the IEEE corpus (IEEE, 1969) were recorded by ten talkers (four male, four female, two children). One block was used in each test session. All the sentences were different. There were five key words in each sentence. Subjects were asked to repeat all the words they heard, and the experimenter recorded which key words had been identified correctly.

C. Consonant test

Twenty /ɑː/-consonant-/ɑː/ nonsense syllables were included, incorporating the consonants /b tʃ d f g h dʒ k l m n p r s ʃ t θ v w z/. Presentation was computer controlled. Each consonant was displayed orthographically on a computer touch screen using its usual spelling (e.g., the sound /tʃ/ was written "CH"). Subjects reported the consonant in each stimulus by touching its orthographic transcription. There were 200 trials in each test. In experiments 1, 2, and 4, materials were recorded by five talkers (two male, two female, one child), and each talker recorded two tokens of each syllable. In experiments 3 and 5, materials were recorded by ten talkers (four male, four female, two children), and each talker recorded a single token of each syllable.

D. Vowel test

Ten h-vowel-d words were included, containing five short vowels: /æ/ (had), /e/ (head), /ɪ/ (hid), /ɒ/ (hod), /ʊ/ (hood), and five long vowels: /ɑː/ (hard), /ɜː/ (heard), /iː/ (heed), /ɔː/ (hoard), /uː/ (who'd). Presentation was computer controlled. Each word was displayed orthographically on a computer touch screen. Subjects responded by touching the orthographic transcription of the appropriate word. There were 200 trials in each test. In experiments 1, 2, and 4, materials were recorded by five talkers (two male, two female, one child). Each talker recorded two tokens of each word, which were presented twice. In experiments 3 and 5, materials were recorded by ten talkers (four male, four female, two children). Each talker recorded two tokens of each word.

E. Speech processing

Speech processing was performed in real time with an eight-channel noise-excited vocoder (Shannon et al., 1995) implemented on a SHARC digital processor (Analog Devices ADSP21065L). Following Rosen et al. (1999), speech signals were analyzed with sixth-order elliptical IIR filters with center frequencies of 433, 642, 925, 1306, 1820, 2513, 3449, and 4712 Hz. Filtered signals were half-wave rectified and low-pass filtered at 160 Hz. The resulting waveform envelopes were multiplied by a white noise that had been low-pass filtered at 10 kHz. The resulting signal in each channel was then filtered by a sixth-order elliptical IIR filter whose center frequency had been shifted relative to the analysis filter in that channel in accordance with Greenwood's (1990) place-to-frequency function to simulate either a 3- or a 6-mm tonotopic shift. When a 6-mm shift was simulated (in experiments 1, 3, 4, and 5), the center frequencies of the reconstruction filters were 1206, 1685, 2332, 3205, 4382, 5971, 8115, and 11 007 Hz. When a 3-mm shift was simulated (in experiment 2), the center frequencies were 741, 1057, 1485, 2061, 2839, 3889, 5305, and 7216 Hz.

F. Subjects

All subjects had normal hearing (≤25 dB HL at octave frequencies between 250 and 8000 Hz, inclusive) measured according to British Society of Audiology (BSA) guidelines (BSA, 1981). All subjects were native speakers of British English and were aged between 18 and 53 years (median age 20 years). Subjects were students or staff of the Universities of Nottingham and York. None of the subjects took part in more than one experiment.

III. EXPERIMENT 1

This experiment investigated the effectiveness of a word-based training task in improving the ability to identify

TABLE I. Design of experiments 1, 2, and 3.

Group	Day 1		Day 2	Day 3	
1	Test session 1	High-variability auditory training	Test session 2	Visual control training	Test session 3
2	Test session 1	Visual control training	Test session 2	High-variability auditory training	Test session 3
3	Test session 1	Single-talker auditory training	Test session 2	Visual control training	Test session 3
4	Test session 1	Visual control training	Test session 2	Single-talker auditory training	Test session 3

words in sentences, vowel sounds, and consonant sounds. Improvements in speech perception following auditory training were compared with improvements following a matched visual control task. The effectiveness of high-variability auditory training, in which training materials were recorded by ten different talkers, was compared with the effectiveness of single-talker auditory training.

A. Method

1. Subjects and speech processing

Sixteen subjects listened to speech through a noise-excited vocoder, which simulated a 6-mm tonotopic misalignment.

2. Training and control tasks

Training was provided by a two-alternative forced-choice task. At the start of each trial, two words were presented orthographically on the left and right of a computer touch screen. The target word was then presented. Subjects responded by touching the word corresponding to the target. Visual feedback on accuracy was given, with a green check indicating that the subject had responded correctly, and a red cross indicating that an incorrect decision had been made. If subjects were incorrect, the trial was repeated until the correct response was given. During auditory training, the target was presented acoustically. During the control task, the target appeared orthographically in the center of the screen degraded by visual noise.

To construct the training materials, 200 key words were selected from 40 IEEE sentences. Three foils were created for each key word, forming quasi-minimal pairs. Over the course of 1200 trials, each key word was presented as the target word itself with each of its three foils, and the three foils were presented as the target word with the key word as the alternative. For high-variability auditory training, ten talkers (four male, four female, two children) recorded the 800 words, with each talker recording 80 words. A single male talker with a southern British accent recorded all 800 words for the single-talker condition. The control task was the same for both the high-variability and single-talker conditions. The auditory training and visual control tasks took approximately 1 h to complete.

3. Design and procedure

Four groups of four participants participated in three sessions. Groups 1 and 2 received high-variability training.

Groups 3 and 4 received single-talker training. Groups 1 and 3 received auditory training between test sessions 1 and 2, while groups 2 and 4 received auditory training between test sessions 2 and 3 (Table I). Sessions took place on consecutive days. During the first session, subjects completed baseline tests of speech perception (test session 1). During the second session subjects completed the auditory training task or the visual control task, followed by further tests of speech perception (test session 2). In the final session, subjects completed either the control task or the auditory training task, again followed by further tests of speech perception (test session 3).

4. Analyses

Analyses were based on changes in performance, measured in percentage points (%pts)[1] between adjacent test sessions. We distinguished changes associated with auditory training from changes associated with visual control training. For groups 1 and 3, the change following auditory training was the difference in score between test sessions 1 and 2, and the change following control training was the difference in score between test sessions 2 and 3. For groups 2 and 4, the change following auditory training was the difference in score between test sessions 2 and 3, and the change following control training was the difference in score between test sessions 1 and 2. The first analysis tested whether either auditory or control training led to a significant improvement in performance. One-sample t tests were performed on the changes in performance following auditory and, separately, control training. The second analysis tested whether auditory training was more effective than the control task in improving speech perception. This hypothesis was tested with paired-samples t tests. The final analysis tested whether high-variability auditory training led to larger improvements in performance than single-talker auditory training. This hypothesis was tested with independent samples t tests on the change following auditory training according to whether subjects received high-variability or single-talker training.

B. Results

1. IEEE Sentence test

Figure 1 (panel a) shows the mean change, and the spread among subjects, in the ability to identify words in IEEE sentences following auditory and control training. The mean improvements and 95% confidence intervals are shown in Table II. There was a significant improvement in sentence
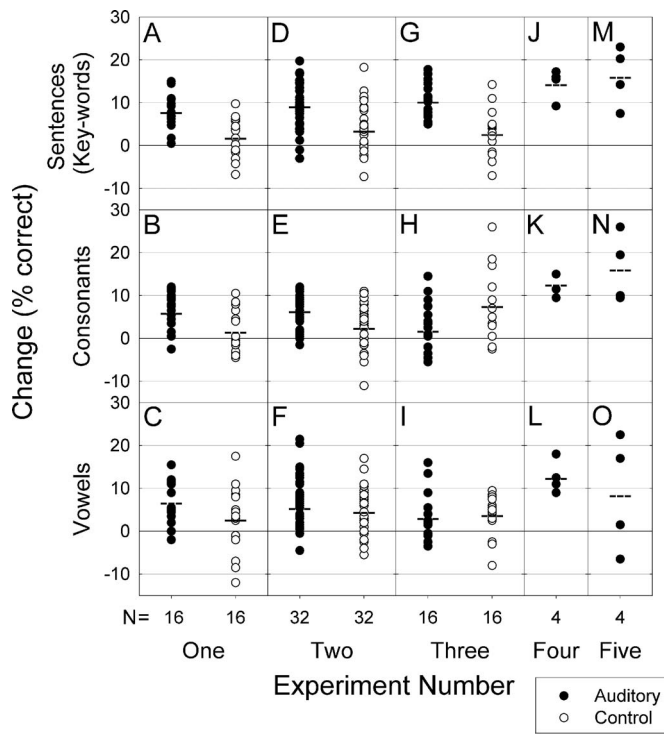
FIG. 1. Results of experiments 1–5. Changes in accuracy of identifying key words, consonants, and vowels by individual subjects. Improvements following auditory training are plotted as filled circles and improvements following the control task are plotted as open circles. The average improvement of each group is indicated by the dashed line. The numbers below the abscissa indicate the number of subjects in each group.

FIG. 2. Results of experiment 1: Percentage of key words correctly identified in IEEE sentences according to test session and training group. The mean value is represented by the dashed line in the box, the median by the solid line. The box spans the interquartile range. Outliers are plotted as dots beyond the 10th–90th percentile whiskers. Groups 1 and 3 received auditory training between test sessions 1 and 2 and visual control training between test sessions 2 and 3. The order was reversed for groups 2 and 4.

perception following auditory training ($t_{15}=8.09$, $p<0.001$), but not following control training ($t_{15}=1.81$). In addition, there was a significantly larger improvement following auditory training than following the control task ($t_{15}=3.13$, $p<0.01$). There was no significant difference between the effectiveness of high-variability and single-talker auditory

training ($t_{14}=0.83$). Figure 2 summarizes these effects by plotting the percentages of key words identified correctly in IEEE sentences according to test session, with data collapsed over variability.

Five of the talkers who recorded the test of sentence perception also recorded the training materials ("old" talkers) and five did not ("new" talkers). There were no significant differences between the effectiveness of high-variability and single-talker auditory training when talkers were "old" ($t_{14}=1.10$) or "new" ($t_{14}=0.14$).

2. Consonant test

There was a significant improvement in sentence perception following auditory training ($t_{15}=5.54$, $p<0.001$), but

TABLE II. Mean (and 95% confidence intervals) improvements (in %pts) in the auditory training task (auditory) and the visual control task (control), and in the high-variability and single-talker versions of the auditory training task.

		Auditory	Control	High-variability auditory	Single-talker auditory
Experiment 1	Sentences	7.98 (5.88 to 10.09)	2.02 (−0.36 to 4.39)	7.16 (3.21 to 11.10)	8.81 (6.23 to 11.40)
	Consonants	6.16 (3.79 to 8.52)	1.75 (−0.86 to 4.36)	7.13 (3.59 to 10.66)	5.19 (1.23 to 9.14)
	Vowels	6.91 (3.72 to 10.09)	2.88 (−1.20 to 6.95)	7.06 (2.80 to 11.33)	6.75 (0.81 to 12.69)
Experiment 2	Sentences	9.38 (7.41 to 11.36)	3.63 (1.61 to 5.65)	11.53 (8.67 to 14.40)	7.23 (4.67 to 9.80)
	Consonants	6.02 (4.59 to 7.44)	2.27 (0.43 to 4.10)	7.28 (5.20 to 9.36)	4.75 (2.79 to 6.71)
	Vowels	7.05 (4.87 to 9.23)	4.17 (2.20 to 6.14)	6.66 (2.44 to 10.87)	7.44 (5.54 to 9.33)
Experiment 3	Sentences	10.44 (8.21 to 12.67)	2.89 (−0.02 to 5.80)	8.63 (4.97 to 12.28)	12.25 (9.50 to 15.00)
	Consonants	2.03 (−1.33 to 5.39)	7.75 (3.61 to 11.89)	3.38 (−2.88 to 9.63)	0.69 (−3.49 to 4.87)
	Vowels	3.25 (0.35 to 6.15)	3.97 (1.43 to 6.51)	4.00 (−0.78 to 8.78)	2.50 (−2.04 to 7.04)

not following control training [$t_{15}=1.43$; Fig. 1(b); Table II]. However, the difference between the change in performance following the auditory training task compared with the control task just failed to reach significance at the <0.05 level ($t_{15}=2.06$, $p=0.057$). There was no significant difference between the effectiveness of high-variability and single-talker auditory training ($t_{14}=0.86$; Table II).

3. Vowel test

There was a significant improvement in vowel discrimination following auditory training ($t_{15}=4.62$, $p<0.001$), but not following control training [$t_{15}=1.50$; Fig. 1(c), Table II]. There was no significant difference between the change in performance following auditory training compared with the control task ($t_{15}=1.31$). There was no significant difference between the effectiveness of high-variability and single-talker auditory training ($t_{14}=1.10$; Table II).

C. Discussion

These results confirm previous demonstrations (Rosen *et al.*, 1999; Fu *et al.*, 2005b; Faulkner *et al.*, 2006) that auditory training improves the ability to perceive spectrally distorted speech. A computer-based word training task, similar to that used by Fu *et al.* (2005b), was associated with significant improvements in the ability to identify words in sentences. Moreover, improvements in the ability to identify words in sentences were significantly larger following auditory training than following a visual control task that exposed subjects to the same vocabulary and imposed the same task demands as the auditory training task. By subtracting the improvement following control training (2.0 %pts) from the improvement following auditory training (7.9 %pts), an improvement of 5.9 %pts in sentence perception can be attributed to perceptual learning resulting from auditory training.

Auditory training was also associated with significant improvements in consonant and vowel discrimination, whereas control training was not. Previously, Fu *et al.* (2005b) reported that word-based training led to significant improvements in consonant and vowel discrimination, while no significant improvements were found for a control group. However, the present experiment did not find that auditory training was associated with significantly *larger* improvements in consonant or vowel recognition than control training. Thus, the effect of auditory training on consonant and vowel discrimination was weaker than its effect on the ability to identify words in sentences. In contrast to these results, Fu *et al.* (2005b) reported a significantly larger improvement in consonant discrimination for a group who received word-based training than for a control group who undertook testing but not training. Part of the difference between the present result and Fu's may have arisen because we controlled for exposure to vocabulary and attention/arousal, along with exposure to test materials. Fu *et al.* (2005b) did not report whether the improvement for the word-based training group was significantly larger than the improvement for the control group on the test of vowel discrimination.

There was no evidence that high-variability auditory training was more effective than single-talker auditory training, and high-variability training did not lead to greater transfer to new talkers. An explanation for this difference may be found in the processing that generates noise-vocoded speech, which strips away many of the cues that distinguish talkers (Gonzalez and Oliver, 2005; Chinchilla-Rodriguez *et al.*, 2004). Gonzalez and Oliver (2005) examined the ability to identify talker gender and speaker identity in noise-vocoded speech. With a similar noise-band vocoder as used in the present study (eight-channel, 160-Hz low-pass filtering of the envelopes within channels), subjects identified the gender of speakers with 89% accuracy and could identify speakers with 78% to 84% accuracy. Thus, although discrimination of different speakers is possible with noise-vocoded speech, performance is far from perfect. In addition, we would expect poorer differentiation between talkers given the size of the tonotopic misalignment that we simulated. Chinchilla-Rodriguez *et al.* (2004) reported that 32 channels were required to discriminate voice gender when speech was spectrally shifted by an octave (to simulate a tonotopic misalignment between 4 and 5 mm), compared with 16 channels for unshifted speech. It is possible, therefore, that differences between the effectiveness of training according to variability would be found if more cues that differentiate different talkers were retained by, for example, simulating a smaller degree of tonotopic misalignment.

IV. EXPERIMENT 2

Average levels of performance in experiment 1 were low. Possibly, the effects of training would be larger if the amount of tonotopic misalignment was reduced so that subjects were operating on a steeper part of the psychometric function relating tonotopic misalignment to performance (Fu and Shannon, 1999). Experiment 2 examined the effectiveness of high-variability and single-talker auditory training when a 3-mm tonotopic misalignment was simulated.

A. Method

The methods were the same as in experiment 1, with the following exceptions. Thirty-two volunteers were tested, with eight subjects in each group (Table I). A tonotopic misalignment of 3 mm was simulated. The test sentences used in test sessions 2 and 3 were counterbalanced across subjects.

B. Results

1. IEEE sentence test

Figure 3 shows the percentages of key words in IEEE sentences reported correctly according to test session for groups 1 and 3 combined and groups 2 and 4 combined. On average, 50.1% of key words were identified correctly at baseline. Both auditory ($t_{31}=9.70$, $p<0.001$) and control training ($t_{31}=3.67$, $p<0.001$) were associated with significant improvements in the ability to identify words in sentences [Fig. 1(d); Table II]. Auditory training led to significantly larger improvements than control training ($t_{31}=3.38$, $p<0.01$). High-variability auditory training was associated with a significantly larger improvement in the percentage of
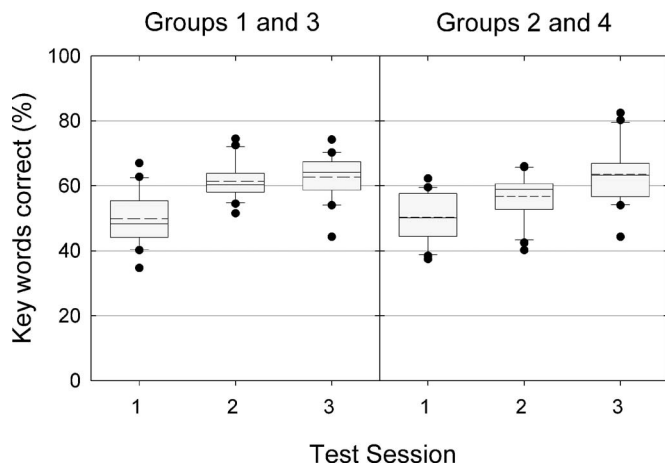
FIG. 3. Results of experiment 2: Percentage of key words correctly identified in IEEE sentences according to test session and training group.

key words correctly identified than single-talker auditory training ($t_{30}=2.38$, $p<0.05$, Fig. 4; Table II).

An independent samples t test on the improvement in accuracy of identifying words spoken by "old" talkers revealed no significant difference between the effectiveness of high-variability (mean improvement=9.9 %pts, 95% c.i., 5.7 to 14.1 %pts) and single-talker (mean improvement =5.9 %pts, 95% c.i., 0.8 to 10.9 %pts; $t_{30}=1.31$, $p=0.20$) auditory training. There was some evidence of a difference between the effectiveness of high-variability (mean improvement=13.1 %pts, 95% c.i., 10.0 to 16.2 %pts) compared with single-talker (mean improvement=8.6 %pts, 95% c.i., 5.0 to 12.3 %pts) auditory training when talkers were "new" ($t_{30}=2.03$, $p=0.051$).

2. Consonant test

There were significant improvements in consonant discrimination following both auditory ($t_{31}=8.63$, $p<0.001$) and control training [$t_{31}=2.52$; Fig. 1(e); Table II]. However, auditory training led to significantly larger improvements
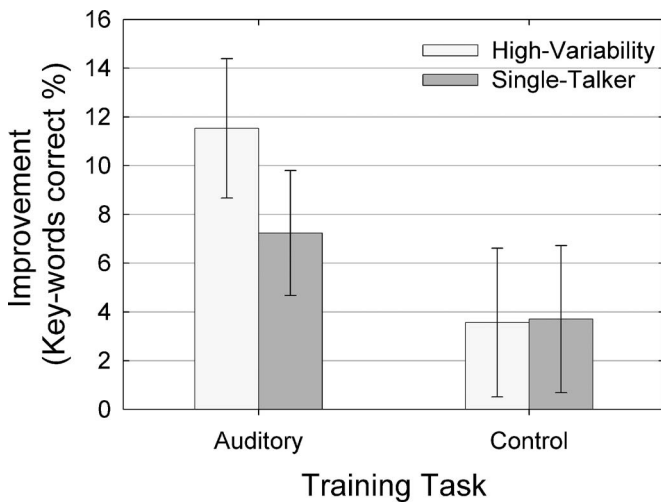
FIG. 4. Results of experiment 2: Improvement in the percentage of key words correctly identified following the auditory training and control tasks according to whether subjects received high-variability or single-talker auditory training. Error bars indicate 95% confidence intervals.

than control training ($t_{31}=2.69$, $p<0.05$). The difference between the effectiveness of high-variability and single-talker auditory training just failed to reach significance ($t_{30}=1.89$, $p=0.069$; Table II).

3. Vowel test

Vowel discrimination improved significantly following auditory ($t_{31}=6.60$, $p<0.001$) and control training [$t_{31}=4.31$, $p<0.001$; Fig. 1(f); Table II]. There was no significant difference between the effectiveness of auditory and control training ($t_{31}=1.63$). The improvements following high-variability and single-talker auditory training did not differ significantly ($t_{30}=0.36$, $p=0.72$; Table II).

C. Discussion

Experiments 1 and 2 show that training-related improvements in performance occur both when baseline performance is poor ($\approx 10\%$, experiment 1) and better ($\approx 50\%$, experiment 2). Perceptual learning in experiment 2 accounted for a significant improvement of 5.8 %pts in ability to identify words in sentences, compared with 5.9 %pts in experiment 1. In addition, experiment 2 found that auditory training led to a significantly larger improvement in the ability to discriminate consonant sounds than control training. As in experiment 1, auditory training was not significantly more effective than the control task in improving the ability to discriminate vowel sounds.

In contrast to experiment 1, experiment 2 found an advantage for high-variability training over single-talker auditory training. The high-variability auditory training task produced an improvement of 12 %pts in the ability to perceive words in sentences, compared with an improvement of 7 %pts following single-talker training. The advantage for high-variability over single-talker auditory training was stronger when talkers were "new" than when talkers were "old". This result is compatible with earlier findings (Lively *et al.*, 1993) that high-variability auditory training can lead to greater transfer to novel talkers than single-talker training. It is possible that this result emerged in experiment 2 but not in experiment 1 because the degree of tonotopic misalignment simulated in experiment 2 preserved more cues that distinguish talkers.

V. EXPERIMENT 3

Experiment 3 sought to establish whether larger improvements could be achieved with a different training task, which required subjects to discriminate words in sentences. We reasoned that a sentence training task might improve performance on a sentence test more than a word training task because there is evidence that auditory training generalizes best when training and test materials are similar (Greenspan *et al.*, 1988; Hirata, 2004).

We retained the comparison between high-variability and single-talker training because we expected there to be larger differences between talkers when they articulated entire sentences compared with single words. Accordingly, we hypothesized that a sentence training task including several talkers would be more effective than a task including only

one talker. In experiments 1 and 2, improvements in speech perception were assessed immediately following auditory training. In the present experiment, an additional testing session was included, approximately 2 weeks after the final training session, to establish whether learning was sustained over time. A spectral shift of 6 mm rather than 3 mm was simulated to avoid possible ceiling effects with male talkers, given that performance with the most intelligible male talker reached 91% key words correct in experiment 2.

A. Method

1. Subjects and speech processing

Sixteen volunteers took part. A 6-mm tonotopic misalignment was simulated.

2. Training materials

Each trial of the auditory training task began with an acoustic presentation of the target sentence. Six orthographically presented words then appeared in random positions on the computer screen. Subjects were instructed to select the three words from this set that were present in the target sentence. Visual feedback on accuracy was given, with a green check indicating that the subject had selected a word that was in the sentence, and a red cross indicating that the subject had selected a word that was not present. If subjects selected a word that was not in the sentence, the sentence was presented again acoustically. Once all three target words had been selected, the target sentence was displayed orthographically at the top of the screen. Subjects were asked to study the sentence. Finally, the sentence was presented acoustically once more. Subjects were asked to listen carefully to the sentence and attempt to pick out words in the sentence that they now knew were present. The aim was to maximize the amount of lexical feedback that subjects received. This protocol is analogous to the distorted-clear-distorted (DCD) protocol that was found to maximize learning to perceive noise-vocoded speech by Davis *et al.* (2005). Our implementation includes an additional intervening task that allows performance to be monitored and that maintains subjects' engagement. The control task was presented in the same format, except that the target sentences were presented orthographically, degraded by visual noise. The training and control tasks took between 1.5 and 2 h to complete.

Three-hundred IEEE sentences that were not used as testing materials were selected as training sentences. Three words in each sentence were selected to be target words. We selected target words that were not highly semantically related, for example, in the sentence "He wrote his last novel at this inn.," "wrote," "last," and "inn" were selected as target words, thus avoiding the semantically related words "wrote" and "novel" both being selected as targets. One foil was created for each target word, so as to form a quasi-minimal pair with the target (e.g., "note," "list," and "it" were selected as foils for the sentence above). For the high-variability training condition, ten talkers (four male, four female, two children) recorded the 300 sentences, with each
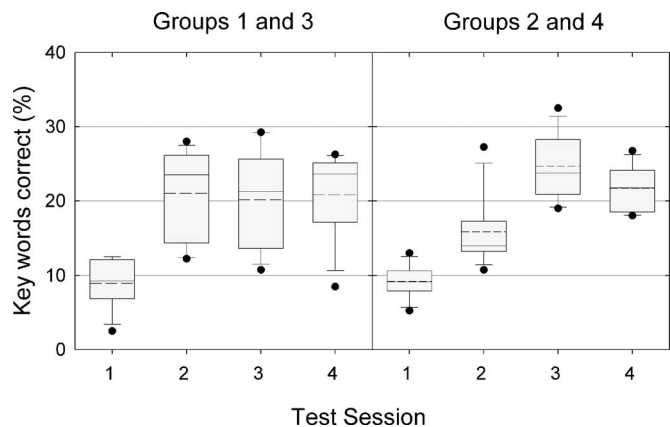
FIG. 5. Results of experiment 3: Percentage of key words in IEEE sentences correctly identified according to test session and training group.

talker recording 30 sentences. A single male talker with a southern British accent recorded all 300 sentences for the single-talker condition.

3. Design, procedure, and analysis

There were four groups of subjects, with four subjects in each group. The design and procedure were the same as in experiment 1 (Table I), but with the addition of a fourth testing session (test session 4) that took place 9–18 days (median 13 days) after test session 3. In this final session subjects received no training. They just completed the tests of speech perception. The results were analyzed in the same way as in experiment 1.

B. Results

1. IEEE sentence test

Figure 5 shows the percentage of key words correctly identified in test sessions 1–4, according to whether subjects received auditory training or control training first. The figure suggests that improvements in performance followed auditory training, and that the level of performance was sustained for both groups over the 2-week interval between test sessions 3 and 4. There was a significant improvement following auditory training ($t_{15}=9.98$, $p<0.001$), while the improvement following control training just failed to reach significance at the <0.05 level [$t_{15}=2.12$, $p=0.051$; Fig. 1(g); Table II]. The improvement following auditory training was significantly larger than the improvement following control training ($t_{15}=3.64$, $p<0.01$). There was no significant difference between the effectiveness of high-variability and single-talker auditory training ($t_{14}=-1.87$; Table II).

Five of the ten talkers who recorded the test of sentence perception also recorded the training materials, giving five "new" talkers and five "old" talkers. There was no evidence that transfer of learning differed between the high-variability and single-talker groups according to whether performance with new ($t_{15}=-1.74$) or old talkers ($t_{15}=-1.31$) was analyzed.

P. C. Stacey and A. Q. Summerfield: Computer-based auditory training

2. Consonant test

Auditory training was not associated with an improvement in the ability to discriminate consonant sounds (t_{15} =1.29), but control training was [t_{15}=3.99, $p<0.01$; Fig. 1(h); Table II]. Control training was not significantly more effective than auditory training, however (t_{15}=1.86). There was no significant difference between the effectiveness of high-variability and single-talker auditory training (t_{14} =0.85; Table II).

3. Vowel test

Significant improvements in vowel discrimination followed both auditory (t_{15}=2.39, $p<0.05$) and control training [t_{15}=3.33, $p<0.01$; Fig. 1 (i); Table II]. There was no significant difference between auditory and control training (t_{15}=−0.31), and no significant difference between high-variability and single-talker auditory training (t_{14}=0.54; Table II).

C. Discussion

The results are partly consistent with those of experiments 1 and 2. Sentence-based auditory training was significantly more effective than a visual control task in improving the ability to identify words in sentences. The average improvement of 7.5 %pts that could be attributed to perceptual learning was comparable to the average improvements of 5.9 and 5.8 %pts found in experiments 1 and 2, respectively. Improvements in sentence perception were sustained over a 2-week period during which no additional training was provided. Changes in mappings between acoustic-phonetic information and linguistic knowledge were therefore "relatively long-lasting"—a requirement for perceptual learning to have occurred (Goldstone, 1998).

In contrast with experiment 2, auditory training was no more effective than control training in improving the discrimination of consonants in nonsense syllables. This result may have occurred because the word-based training regime used in experiment 2 provides more directed training in the discrimination of consonant sounds, being based on quasi-minimal pairs of words, the majority of which differed in terms of the consonant sound.

High-variability auditory training was not more effective than single-talker training. As discussed in experiment 1, this result may have arisen because the 6-mm tonotopic misalignment was too severe to permit differentiation between talkers.

VI. EXPERIMENT 4

Experiments 1 and 2 demonstrated that word-based auditory training was more effective than a visual control task in improving the ability to identify words in sentences and consonants in nonsense syllables. Experiment 4 examined whether performance continued to improve if subjects repeated the word-based high-variability auditory training task four times, for a total of approximately 4 h of training.
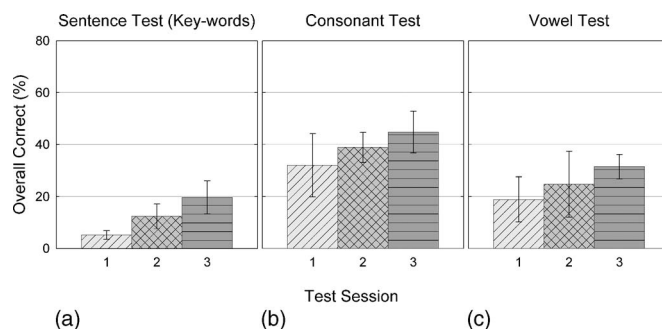
FIG. 6. Results of experiment 4: Overall performance on the sentence test (panel a), the consonant test (panel b), and the vowel test (panel c) in test sessions 1, 2, and 3.

A. Method

Four volunteers participated. They each completed four sessions of high-variability word-based training, providing a total of 4800 training trials. As in experiment 1, during test session 1, subjects completed baseline tests of speech perception. Test session 2 took place on the following day, and subjects completed the auditory training task followed by further tests of speech perception. On subjects' third and fourth visits they repeated the training task. In their fifth visit they completed the training task followed by further tests of speech perception (test session 3). The third to fifth visits were scheduled to take place at any time within a 2-week period. The test sentences used in test sessions 2 and 3 were counterbalanced across subjects. Results were analyzed using repeated measures analyses of variance. A Greenhouse-Geisser correction was applied if the assumption of sphericity was violated (indicated by noninteger degrees of freedom). Planned comparisons were carried out using t tests with a Bonferroni correction.

B. Results

1. IEEE sentence test

Performance improved significantly over test sessions 1–3 [$F_{1.0,3.1}$=60.32, $p<0.01$; Fig. 1(j)]. In test session 1, subjects correctly reported 5.3% of words in sentences. There was a significant improvement of 7.1 %pts following one training run between test sessions 1 and 2 (planned t_3 =5.34, $p<0.05$) and of 7.4 %pts following three further training runs between test sessions 2 and 3 [planned t_3 =14.60, $p<0.001$; Fig. 6(a)].

2. Consonant test

Performance on the consonant test also improved significantly over sessions [$F_{2,6}$=25.25, $p<0.01$; Fig. 1(k)]. Initially, subjects identified 32.0% of consonant sounds correctly. There was an improvement in performance of 6.9 %pts between test sessions 1 and 2, which failed to reach significance (planned t_3=2.89), and an improvement of 5.9 %pts between test sessions 2 and 3, which just failed to reach significance [planned t_3=3.98, $p=0.056$; Fig. 6(b)]. There was a significant improvement of 12.8 %pts between test sessions 1 and 3 (planned t_3=9.36, $p<0.01$).
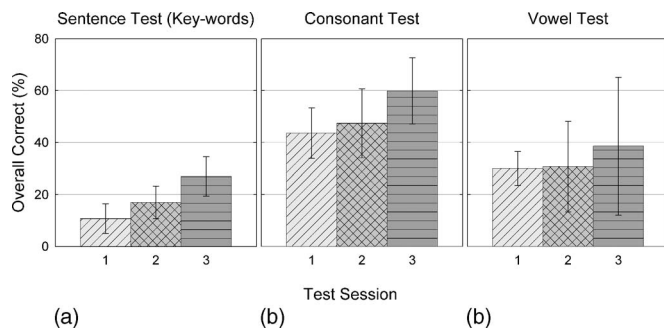
FIG. 7. Results of experiment 5: Overall performance on the sentence test (Panel a), the consonant test (Panel b) and the vowel test (Panel c) in test Sessions 1, 2, and 3.

3. Vowel test

The ability to identify vowel sounds improved over sessions [$F_{2,6}=14.08$, $p<0.01$; Fig. 5(c)]. There was a significant improvement of 12.6 %pts between test sessions 1 and 3 [planned $t_3=6.50$, $p<0.05$, Fig. 1(l)], but no significant improvements between adjacent test sessions.

C. Discussion

Experiment 4 established that performance continues to improve if additional auditory high-variability training sessions are provided following an initial 1-h session. Accuracy of identification of words in sentences improved from 5.3% to 19.8% correct after four training sessions. The improvement of 7.1 %pts following the first training session is comparable to the improvement of 6.3 %pts displayed by the equivalent group of participants in experiment 1 who received a single session of high-variability auditory training between test sessions 1 and 2. A similar pattern was found in the tests of consonant and vowel discrimination. Improvements after one training session were similar to those found in experiment 1. The improvements roughly doubled following three more training sessions.

VII. EXPERIMENT 5

The aim of experiment 5 was to establish whether additional sentence-based training sessions would produce further improvements after the initial session. As in experiment 4, subjects repeated the high-variability auditory training task four times, for a total of 6–7 h of training.

A. Method

The methods were the same as those described in experiment 4, except that the sentence-based training task was administered. Subjects completed a total of 1200 training trials, with the same 300 sentences being presented during each of the four training sessions.

B. Results

1. IEEE sentence test

A one-way ANOVA revealed that performance differed between test sessions 1–3 [$F_{2,6}=17.07$, $p<0.001$; Fig. 7(a)].

Performance improved from 10.8% correct in test session 1 to 27.0% correct in test session 3. There was an overall significant (planned $t_3=4.72$, $p<0.05$) improvement of 16.3 %pts between test sessions 1 and 3 [Fig. 1(m)], but no significant improvement from test sessions 1 to 2, or 2 to 3.

2. Consonant test

Performance differed significantly between test sessions ($F_{2,6}=13.16$, $p<0.01$). Performance improved from 43.6% correct in test session 1 to 59.9% correct in test session 3. After adjusting for multiple comparisons, there was no significant improvement between adjacent test sessions, although there was a significant improvement of 16.3 %pts between test sessions 1 and 3 [planned $t_3=4.08$, $p<0.05$; Fig. 1(n)].

3. Vowel test

The percentage of vowel sounds identified correctly improved from 30.0% in test session 1 to 38.6% in test session 3. A one-way ANOVA on the performance in test sessions 1–3 did not reveal a significant main effect ($F_{2,6}=1.75$). There were no significant improvements between adjacent test sessions, and no overall significant improvement between test sessions 1 and 3 [Fig. 1(o)].

C. Discussion

Performance on the IEEE sentence test continued to improve with successive training sessions. The average improvement in the ability to identify words in sentences was 16.3 %pts, with a range of improvements across subjects of 7.5 to 23.0 %pts. The improvement of 23 %pts is the largest found in any of the present series of experiments and approach the gains found in other investigations that have used one-to-one training (e.g., Rosen *et al.*, 1999; Faulkner *et al.*, 2006). There was additionally a significant improvement of 16.3 %pts in consonant discrimination by the end of training, but no significant improvement in vowel discrimination. As experiment 3 found no significant difference between the effectiveness of sentence-based auditory training compared with the control task on the test of consonant discrimination, we cannot be confident that the improvement in consonant discrimination emerged because of auditory training, rather than because of repeated testing.

VIII. GENERAL DISCUSSION

The present study examined whether computer-based auditory training could lead to improvements in the ability of normal-hearing listeners to perceive spectrally distorted speech that had been processed by a simulation of a cochlear-implant system. The five experiments have revealed that word- and sentence-based approaches to training lead to significant improvements in the ability to identify words in sentences, and these improvements are significantly larger than those that follow a visual control task. We have therefore extended Fu *et al.* (2005b) findings by showing that both word- and sentence-based training lead to significant improvements in the ability to identify words in sentences. Ex-

tending the amount of auditory training beyond the initial hour led to continued improvements in the ability to identify words in sentences, ranging between 8 and 23 %pts. These improvements approach those reported by Rosen *et al.* (1999), who provided one-to-one training using CDT, although they are on average somewhat smaller. However, the evidence suggests that performance would have continued to improve if more training had been provided.[2]

A further significant finding was that there was an advantage for high-variability auditory training over single-talker auditory training in experiment 2, which simulated a 3-mm tonotopic misalignment. The results are compatible with the hypothesis (Lively *et al.*, 1993) that high-variability training generalizes better to new talkers than single-talker training. No significant differences between high-variability and single-talker training emerged in experiments 1 and 3, possibly because the simulation of a 6-mm tonotopic misalignment, with eight channels of information, was too extreme to permit subjects to differentiate between talkers (Chinchilla-Rodriguez *et al.*, 2004).

A. Difficulty training vowel discrimination

An advantage for word-based training over the visual control task was found for consonant discrimination, but not for vowel discrimination. It seems, therefore, that vowel discrimination is more difficult to train than consonant discrimination. This might be because the speech processing preserved the predominantly temporal cues required for consonant discrimination to a greater extent than the predominantly spectral cues required for vowel discrimination. Vowel discrimination also depends on the interaction between formant frequencies and the fundamental frequency (f_0; Assman *et al.*, 2002; Assman and Nearey, 2003). Assman and Nearey (2003) reported that vowel discrimination was adversely affected when f_0 remained constant and formants were shifted upward or downward in frequency. The decline in performance was counteracted, however, when corresponding upward or downward shifts in f_0 were introduced. The f_0 information is not strongly preserved in noise-vocoded speech, particularly in the speech of women and children, possibly contributing to the difficulties reported here.

B. Training-related versus incidental learning

In order to dissociate training-related learning from incidental learning, improvements following auditory training were compared with improvements following a "matched" visual control task. The visual control task was matched with the auditory training task in terms of its content and task demands. In addition, the visual stimulus was impoverished somewhat to mirror the loss of spectral information in the auditory stimuli. However, the visual stimulus underwent no further distortion, and the consequences of spectral shift were not recreated in the visual control task. The visual control task might have acted as a more rigorous control condition if the consequences of both the loss of spectral detail and spectral shifting could have been recreated in the visual modality.

The suggestion that the auditory and visual control tasks could have been more fully matched is supported by differences in the level of performance between the auditory training task and the visual control task. Performance reached 98% correct in the visual control task in experiment 1, compared with 90% in the auditory training task. Potentially, differences between the effectiveness of auditory compared with control training were simply due to differences in the difficulty of the training task. However, this suggestion is contradicted by the finding that there was no significant difference between the effectiveness of high-variability and single-talker training in experiment 1, despite the fact that performance averaged 85% correct in the high-variability training task, compared with 94% correct in the single-talker training task.

Although we found that auditory training led to significantly larger improvements in speech perception than the visual control task, "incidental learning" also contributed to improvements in performance. Figures 2, 3, and 5 show that the extent to which performance improved following the control task depended on the order in which the auditory training and visual control tasks were completed. Subjects who were exposed to the auditory training task first displayed a significant improvement between test sessions 1 and 2, but no further significant improvement between test sessions 2 and 3 following control training. In contrast, subjects who were exposed to the control task first displayed an improvement between test sessions 1 and 2, and then a further improvement between test sessions 2 and 3 following auditory training. This pattern of results is compatible with the idea that two processes contributed to the improvements in sentence perception. The first process is related to auditory training, while the second is related to incidental learning. Incidental learning happened early in the experiments, occurring between test sessions 1 and 2, but not between test sessions 2 and 3. Accordingly, both training-related and incidental learning caused the improvements between test sessions 1 and 2 for the group who received auditory training first. For the group who received control training first, it is likely that the improvement between test sessions 1 and 2 can be attributed to incidental learning, and that the improvement between test sessions 2 and 3 can be attributed to auditory training.

The present study sought to dissociate training-related learning from incidental learning in order to measure the effectiveness of our training tasks. If the contribution of incidental learning was large, and the contribution of training-related learning was very small, there would be little point in asking subjects to complete an extensive amount of auditory training. Arguably however, the more interesting distinction is between learning that generalizes across speech tests and translates into improvements in everyday life, versus learning that does not generalize. Given that this is a simulation study, and thereby can afford no advantages in terms of subjects' everyday lives, it was not possible to dissociate these two types of learning within the present experiment. In further work with cochlear-implant users, we are evaluating the effectiveness of auditory training using a questionnaire alongside tests of speech perception. Patients use the ques-

tionnaire to report the extent to which training has benefited them in everyday life. In this way, we shall determine whether training produces learning that generalizes beyond laboratory tasks.

C. Basis of improvements following training

There are three potential explanations for why performance improved following auditory training. First, it is possible that subjects learned to remap the novel auditory sensations onto their existing linguistic knowledge. Cochlear-implant users tend to exhibit quite marked improvements in speech perception during the first few months of implant use, and may continue to improve up to 2 years post implantation (Tyler and Summerfield, 1996; Tyler et al., 1997). Evidence suggests that this improvement is driven primarily by improvements in the ability to map the novel sensations provided by a cochlear implant onto existing linguistic knowledge (Svirsky et al., 2001, 2004). It is possible that a similar process underlies the improvements reported here. The word and sentence training tasks led to similar overall levels of improvement in the ability to identify words in sentences. It therefore seems that subjects abstracted general information about the mapping between acoustic properties and phonetic and/or lexical information. If performance had improved more following sentence training than following word training, this might have suggested that the cognitive skill required for sentence perception had improved, rather than a general improvement in the relationship between acoustic input and existing representations. Second, it is possible that subjects learned to hear differences between similar sounds. Goldstone (1998) describes differentiation as one of the major mechanisms of perceptual learning, explaining "stimuli that were once psychologically fused together become separated. Once separated, discriminations can be made between stimuli that were originally indistinguishable" (p. 596). This mechanism is thought to underlie the improvement in the ability of Japanese listeners to discriminate /r/ and /l/ (e.g., Lively et al., 1993). Third, it is possible that a component of the improvement following training arises from a general improvement in listening skills and in the ability to attend to auditory information. However, if a general improvement in listening skills was solely responsible for improvements following training, we would have expected improvements in all tests of speech perception.

D. Implications for cochlear-implant users

A motivation for this study was to establish whether self-administered computer-based training techniques would improve the ability of normal-hearing subjects to perceive spectrally distorted speech, before evaluating whether the training procedures lead to improvements in speech perception amongst cochlear-implant users. The reasoning was that in order for a training regime to be valuable for cochlear-implant users, it should improve the ability of normal-hearing subjects to perceive speech that is reduced in spectral detail and is frequency shifted. The results indicate that our training procedures are successful in improving speech perception in normal-hearing subjects, and they also suggest

that training materials should be recorded by several talkers. However, these results do not necessarily mean that the training procedures will improve speech perception amongst implantees. While cochlear-implant users listen to the distorted input during everyday communication and have the opportunity to learn the relationship between motor speech activity and the resulting auditory sensations, the normal-hearing listeners in our experiments only ever heard the distorted signal within the laboratory and they never hear their own speech through the simulation. Accordingly, the demonstration of learning with normal-hearing subjects can be considered a necessary, but insufficient, condition for auditory training to be successful with cochlear-implant users. However, despite this qualification, similar auditory training approaches have been associated with improved outcomes amongst adult users of cochlear implants (Fu et al., 2005a).

E. Conclusion

The present experiments show that self-administered computer-based auditory training regimes lead to improvements in the ability of normal-hearing listeners to perceive spectrally distorted speech, even when exposure to the training vocabulary and test materials is controlled. The results suggest that the word- and sentence-based training tasks have the potential to improve the perception of speech by adult users of cochlear implants and also suggest that training packages should contain materials spoken by multiple talkers. We are currently investigating whether these training materials lead to benefits for implantees.

[1]The size of changes in %-correct observed following training may depend on baseline performance levels, meaning that an improvement of 20 %pts from 0% correct should not be equated to an improvement of 20 %pts from 40% correct. Three transforms may reduce this problem: (1) differences in arcsine-transformed percentages, (2) percent reduction of error, or (3) calculating the ratio of the logarithms of error probabilities. We applied each of these transforms to our data, but found no difference in the pattern of results.

[2]The relationship between baseline performance and the overall improvement in performance was examined in experiments 1, 2, and 3 for each outcome measure (sentences, consonants, and vowels). Four of the nine correlations were significant, with poorer performers at baseline showing larger improvements. However, there was much variability. Participants at the same baseline level showed improvements ranging from less than 10 %pts to more than 20 %pts, suggesting that baseline performance is only one of the variables that influence the size of the benefit following training.

Amitay, S., Irwin, A., and Moore, D. R. (**2006**). "Discrimination learning induced by training with identical stimuli," Nat. Neurosci. **9**, 1446–1448.
Assman, P. F., and Nearey, T. M. (**2003**). "Frequency shifts and vowel

identification," in *Proc. 15th Int. Congress Phonetic Science*, edited by M. J. Sole, D. Recasens, and J. Romero, Barcelona, Spain, pp. 1–4.

Assman, P. F., Nearey, T. M., and Scott, J. M. (**2002**). "Modelling the perception of frequency-shifted vowels," in *Proc. 7th Int. Cong. Spoken Lang. Proc.*, pp. 425–428.

Baskent, D., and Shannon, R. V. (**2003**). "Speech recognition under conditions of frequency-place compression and expansion," J. Acoust. Soc. Am. **113**, 2064–2076.

Bradlow, A. R., and Bent, T. (**2003**). "Listener adaptation to foreign accented English," in *Proc. 15th Int. Congress Phonetic Science*, edited by M. J. Sole, D. Recasens, and J. Romero, Barcelona, Spain, 2881–2884.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

BSA (**1981**). "Recommended procedures for pure-tone audiometry using a manually operated instrument," Br. J. Audiol. **15**, 213–216.

Busby, P. A., Roberts, S. A., Tong, Y. C., and Clark, G. M. (**1991**). "Results of speech perception and speech production training for three prelingually deaf patients using a multiple-electrode cochlear implant," Br. J. Audiol. **25**, 291–302.

Chinchilla-Rodriguez, S., Nogaki, G., and Fu, Q.-J. (**2004**). "Relative contribution of spectral and temporal cues and spectral profile to voice gender discrimination," J. Acoust. Soc. Am. **116**, 2544–2545.

Clopper, C. G., and Pisoni, D. B. (**2004**). "Effects of talker variability on perceptual learning of dialects," Lang. Speech **47**, 207–239.

Davis, M. H., Hervais-Adelman, A., Taylor, K., McGettigan, C., and Jonsrude, I. S. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Dawson, P. W., and Clark, G. M. (**1997**). "Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant," Ear Hear. **18**, 488–501.

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., Loizou, P. C., and Rainey, D. (**1997**). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," J. Acoust. Soc. Am. **102**, 2993–2996.

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 depth," Ear Hear. **27**, 139–152.

Fu, Q.-J., Galvin, J., 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., and Shannon, R. V. (**1999**). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electrical hearing," J. Acoust. Soc. Am. **105**, 1889–1900.

Gagne, J. P., Parnes, L. S., LaRocque, M., Hassan, R., and Vidas, S. (**1991**). "Effectiveness of an intensive speech perception training program for adult cochlear implant recipients," Ann. Otol. Rhinol. Laryngol. **100**, 700–707.

Gantz, B. J., Woodworth, G. G., Abbas, P. J., Knutson, J. F., and Tyler, R. S. (**1993**). "Multivariate predictors of audiological success with multichannel cochlear implants," Ann. Otol. Rhinol. Laryngol. **102**, 909–916.

Gfeller, K., Witt, S., Ademek, M., Mehr, M., Rogers, J., Stordahl, J., and Ringgenberg, S. (**2002**). "Effects of training on timbre recognition and

appraisal by postlingually deafened cochlear implant recipients," J. Am. Acad. Audiol **13**, 132–145.

Goldstone, R. L. (**1998**). "Perceptual learning," Annu. Rev. Psychol. **49**, 585–612.

Gonzalez, J., and Oliver, J. C. (**2005**). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. **118**, 461–470.

Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (**1988**). "Perceptual learning of speech produced by rule," J. Exp. Psychol. Learn. Mem. Cogn. **14**, 412–433.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species - 29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hirata, J. (**2004**). "Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts," J. Acoust. Soc. Am. **116**, 2384–2394.

IEEE (**1969**). *IEEE Recommended Practice for Speech Quality Measurements* (IEEE, New York).

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variablility in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). "Training Japanese listeners to identify English /r/ and /ll/. III. Long-term retention of new phonetic categories," J. Acoust. Soc. Am. **94**, 2076–2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.

Robinson, K., and Summerfield, A. Q. (**1996**). "Adult auditory learning and training," Ear Hear. **17**, 51S–65S.

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.

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

Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., Neely, J. G., Kletzer, 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.

Summerfield, A. Q., and Marshall, D. H. (**1995**). *Cochlear Implantation in the UK 1990-1994: Main Report* (HMSO Books, London).

Svirsky, M. A., Silveira, A., Suarez, H., Neuburger, H., Lai, T. T., and Simmons, P. M. (**2001**). "Auditory learning and adaptation after cochlear implantation: a preliminary study of discrimination and labeling of vowel sounds by cochlear implant users," Acta Oto-Laryngol. **121**, 262–265.

Svirsky, M. A., Silveria, A., Neuburger, H., Teoh, S.-W., and Suarez, H. (**2004**). "Long-term auditory adaptation to a modified peripheral frequency map," Acta Oto-Laryngol. **124**, 381–386.

Tyler, R. S., Parkinson, A. J., Woodworth, G. G., Lowder, M. W., and Gantz, B. J. (**1997**). "Performance over time of adult patients using the Ineraid or Nucleus cochlear implant," J. Acoust. Soc. Am. **102**, 508–522.

Tyler, R. S., and Summerfield, A. Q. (**1996**). "Cochlear implantation: relationships with research on auditory deprivation and acclimatization," Ear Hear. **17**, 38S–50S.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (**1999**). "Training American listeners to perceive Mandarin tones," J. Acoust. Soc. Am. **106**, 3649–3658.

Yukawa, K., Cohen, L., Blamey, P., Pyman, B., Tungvachirakul, V., and O'Leary, S. (**2004**). "Effects of insertion depth of cochlear implant electrodes upon speech perception," Audiol. Neuro-Otol. **9**, 163–172.

Tone recognition in continuous Cantonese speech using supratone models

Yao Qian and Tan Lee[a]
Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Frank K. Soong
Microsoft Research Asia, 5th Floor, Beijing Sigma Center, 49 Zhichun Road, Haidian, Beijing 100080, China

This paper studies automatic tone recognition in continuous Cantonese speech. Cantonese is a major Chinese dialect that is known for being rich in tones. Tone information serves as a useful knowledge source for automatic speech recognition of Cantonese. Cantonese tone recognition is difficult because the tones have similar shapes of pitch contours. The tones are differentiated mainly by their relative pitch heights. In natural speech, the pitch level of a tone may shift up and down and the F0 ranges of different tones overlap with each other, making them acoustically indistinguishable within the domain of a syllable. Our study shows that the relative pitch heights are largely preserved between neighboring tones. A novel method of supratone modeling is proposed for Cantonese tone recognition. Each supratone model characterizes the F0 contour of two or three tones in succession. The tone sequence of a continuous utterance is formed as an overlapped concatenation of supratone units. The most likely tone sequence is determined under phonological constraints on syllable-tone combinations. The proposed method attains an accuracy of 74.68% in speaker-independent tone recognition experiments. In particular, the confusion among the tones with similar contour shapes is greatly resolved. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717413]

I. INTRODUCTION

Tone refers to the use of different pitch patterns to distinguish the meanings of words in a spoken language (Pike, 1948). In Chinese dialects like Mandarin and Cantonese, each basic lexeme, which corresponds to a written Chinese character, is pronounced as a monosyllable carrying a specific tone. A change of the tone may lead to a change of the meaning. In Cantonese, for example, the syllable *fu*:[1] with different tones can mean "husband (夫)," "tiger (虎)," "rich (富)," "symbol (符)," "woman (婦)," or "father (父)." Given the contrastive lexical function of tone, it is believed that tone information plays an important role in automatic speech recognition (ASR) of Chinese languages. ASR is generally formulated as a computational problem of determining the best word sequence among many possibilities. If the tones are known, the number of candidate words is greatly reduced so that the performance of ASR systems can be improved in terms of both accuracy and efficiency.

Phonologically, tones are described and differentiated by their distinct pitch levels and contours. Acoustically, pitch is measured in terms of the fundamental frequency of the speech signal (henceforth abbreviated as F0). In addition to F0, syllable duration and amplitude contour were found to be useful tone features (Blicher *et al.*, 1990; Whalen and Xu,

1992). Automatic tone recognition is a pattern classification process that determines the tone identities of individual syllables based on these tone features.

Tone recognition of isolated Mandarin syllables was studied by Yang *et al.* (1988), Chang *et al.* (1990), Hon *et al.* (1994), and Lin *et al.* (1996). Most of these works reported very high recognition accuracy. Tone recognition for continuous speech is much more difficult because of the complex contextual variation of F0. Chen and Wang (1995) proposed a method of compensating for the coarticulation effect using neighboring tones and sentential intonation. In Wang *et al.* (1997), context-dependent tone models were employed to deal with the contextual variation of tones. Recently, Zhang *et al.* (2005) proposed the method of tone-nucleus-based modeling, and exploited the pitch contrast between neighboring tones.

The present study focuses on automatic tone recognition in continuous Cantonese speech. Cantonese is a major Chinese dialect that is known for being rich in tones (Hashimoto, 1972; Bauer and Benedict, 1997). In Lee *et al.* (1995), suprasegmental features derived from syllable-wide F0 contours were proved to be effective for tone recognition of isolated Cantonese syllables. In Lee *et al.* (2002a), tone recognition in continuous Cantonese speech was done with context-dependent hidden Markov models (HMM). Similar work was reported on the use of support vector machines for Cantonese tone recognition (Peng and Wang, 2005). Although these studies made good initial attempts, the peculiar linguistic and acoustic properties of Cantonese tones have

---

[a]Electronic mail: tanlee@ee.cuhk.edu.hk

not been fully exploited in the design of tone models. In this paper, we describe a novel method of tone recognition using the so-called supratone models. Each supratone unit covers a number of syllables in succession, and thus a supratone model characterizes not only the F0 contours of individual tones, but also the transitions between them. By including multiple syllables in a single modeling unit, the relative pitch heights of the tones are captured explicitly. This information is useful especially for distinguishing the Cantonese tones that have highly similar shapes of pitch contours.

In Sec. II, a brief description of Cantonese phonology is given and the tone system of Cantonese is described. In Sec. III, acoustic variations of Cantonese tones are discussed and the motivations of supratone modeling are explained. Details of the proposed method for Cantonese tone recognition are given in Sec. IV. Section V presents the experimental results and discussions. Section VI gives our concluding remarks.

II. CANTONESE TONES

Cantonese is used in daily communication by tens of millions of people in the provinces of Guangdong and Guangxi, the neighboring regions of Hong Kong and Macau, and many overseas Chinese communities (Bauer and Benedict, 1997). This research focuses on Cantonese as it is spoken in Hong Kong. All speech materials that are used in the acoustic analysis and recognition experiments were collected from native speakers of Hong Kong Cantonese.

In Cantonese, each Chinese character is pronounced as a tonal syllable. If we consider only the phonemic composition and ignore its tonal variation, the syllable unit is referred to as a *base syllable* (Wang *et al.*, 1997; Lee *et al.*, 2002b). Each base syllable can be divided into two parts: the *initial* or *onset* and the *final* or *rime*. The initial is a consonant. The final typically consists of a vowel nucleus and a consonant coda. The initial and the coda are optional. The simplest syllable consists of a single vowel or a syllabic nasal, and the most complex syllable takes the form "consonant+vowel +consonant." There are 19 initials and 53 finals in Cantonese (Yuan, 1960; Hashimoto, 1972). The initials can be categorized into six classes: Plosives, affricates, fricatives, glides, liquids, and nasals (Ching *et al.*, 1994). The first three classes are voiceless and the last three are voiced. The finals are divided into five categories: vowels (long), diphthongs, vowels with a nasal coda, vowels with a stop coda, and syllabic nasals. The stop codas [p], [t], and [k] are unreleased (Bauer and Benedict, 1997).

Cantonese has been said to have nine or ten tones (Hashimoto, 1972).[2] They are described by different pitch contours, as shown in Fig. 1. A tone contour is a two-dimensional sketch of pitch movement in which the vertical dimension shows the pitch height and the horizontal dimension indicates the length of the tone. In Fig. 1, different systems that have been used to label the tones are compared. They include the historical names of Chinese tone categories (Hashimoto, 1972), pitch contour descriptions (Bauer and Benedict, 1997), the tone letters devised by Chao (1947), and the Jyut Ping system (LSHK, 1997). Cantonese tones are distributed across three pitch levels: high, mid, and low. At
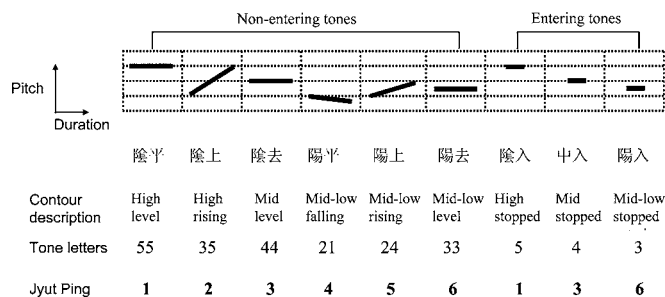
FIG. 1. Pitch patterns of the Cantonese tones and different tone labeling systems.

each level, the tones are further classified according to the shape of the pitch contour. "Entering tone" is a historically defined tonal category that was used as a cover term for tones that co-occur with "checked" syllables, i.e., syllables that end in an occlusive coda such as [p], [t], or [k] (Chen, 2000). Entering tones are contrastively shorter than nonentering tones. In terms of pitch level, each entering tone coincides roughly with one of the nonentering tones. It has been suggested to group the three entering tones with their nonentering counterparts so that the number of tones is reduced to six (Bauer and Benedict, 1997; Cutler and Chen, 1997; Lee *et al.*, 2002c). From acoustic points of view, the duration of tones varies greatly in natural speech, making duration a less reliable feature for tone identification. This research is focused on the problem of recognizing six Cantonese tones without separating the entering tones from the nonentering tones. We use the Jyut Ping system to label the six tones as shown in Fig. 1.

III. ACOUSTIC VARIATION OF CANTONESE TONES

A. Tone contours of isolated syllables and continuous speech

Acoustically, the tone contour of a Cantonese syllable can be visualized via a plot of time-varying F0 over the entire duration of the syllable. In the case that the initial consonant is voiceless, the tone contour covers only the final segment. Throughout this study, the Robust Algorithm for Pitch Tracking (RAPT) described in Talkin (1995) is used for computing F0 from digitized speech signals on a short-time basis. Figure 2 shows the F0 contours of the nine Cantonese tones. They were obtained from 1800 isolated tonal syllables uttered by a male native Cantonese speaker (Lee *et al.*, 2002a). These syllables form a nearly complete syllabary of modern Hong Kong Cantonese. Each contour shown in Fig. 2 is the average over all syllables that carry the respective tone. For ease of comparison, the contours are aligned to the same length. It is observed that the tone contours of isolated syllables reflect the canonical pitch patterns very well. Tones 1, 3, 4, and 6 have either flat or slightly falling F0 contours, and can be viewed as level-pitch tones at different levels. It can also be seen that the length-equalized contours of the entering tones and their nonentering counterparts are very close.

F0 is a highly variable acoustic feature. Apart from the fact that the pitch of a male voice is generally lower than that

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Qian *et al.*: Tone recognition using supratone models   2937
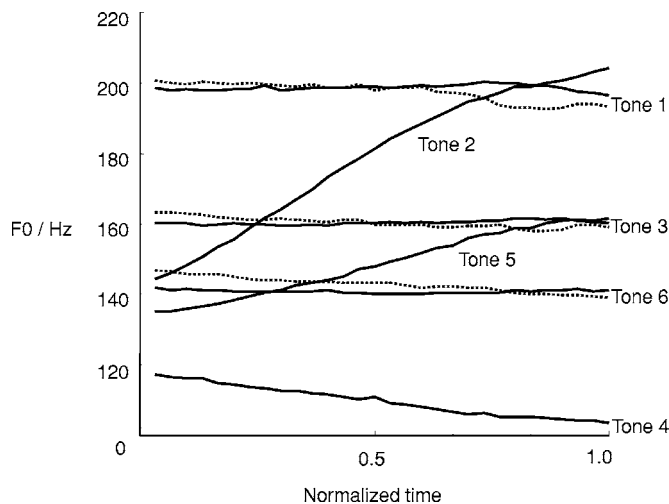
FIG. 2. Average F0 contours of Cantonese tones uttered in isolation (Lee *et al.*, 2002a). The contours of the three entering tones are drawn separately with dashed lines.

FIG. 4. Statistical variation of the F0 contours of the six Cantonese tones uttered by 34 male speakers. Each contour is divided into three sections, and the average logarithmic F0 value of each section is computed. The thick solid bar extends from the 25th to the 75th percentile. The white strip indicates the median. The thin dashed line extends from the 5th to the 95th percentile.

of a female or child voice, F0 is determined by the speaker's physiological characteristics, health condition, speaking style, and emotional status (Xu, 2001). In natural speech, intonation and coarticulation are the major factors that cause tone contours to deviate from their canonical patterns (Xu, 2001; Li *et al.*, 2004; Liu, 2001). Figure 3 shows the waveform and the extracted F0 contour of a Cantonese sentence uttered by a male native speaker. The sentence consists of 19 tonal syllables, which are labeled in Jyut Ping. The numeral at the end of each syllable transcription is the tone marker (1 to 6). For each syllable, an average value of F0 is computed from the middle one-third section of its tone contour, and is shown at the bottom of the F0 plot to represent the pitch level of the respective syllable. It is observed that the F0 contours of individual syllables deviate greatly from their canonical patterns. The sentence contains six syllables of Tone 1 (high level). Their F0 contours exhibit large variation. One major cause is the tone coarticulation, for example, between "heil" and "mong6" (the 14th and 15th syllables). The F0 levels change greatly among syllables with the same tone in the same sentence. For example, the pitch levels of

the six Tone 1 syllables range from 123 to 196 Hz. They seem to be related to the positions of the syllables in the sentence.

B. A statistical analysis of F0 contours

A statistical analysis of tone contours is carried out on a large database of continuous Cantonese speech. This database, named CUSENT, was designed for speaker-independent large vocabulary continuous speech recognition (Lee *et al.*, 2002b). More details of CUSENT can be found in Sec. V A. About 10 000 utterances from 34 male speakers are processed. Each utterance contains a complete sentence of 6–20 syllables. The F0 contour of each syllable is divided evenly into three sections. The average logarithmic F0 value of each section is computed. Figure 4 shows the statistical variation in the six tones. There exists large overlapping among different tones. The confusion is particularly severe between Tone 3 and Tone 6, and between Tone 6 and Tone 4.
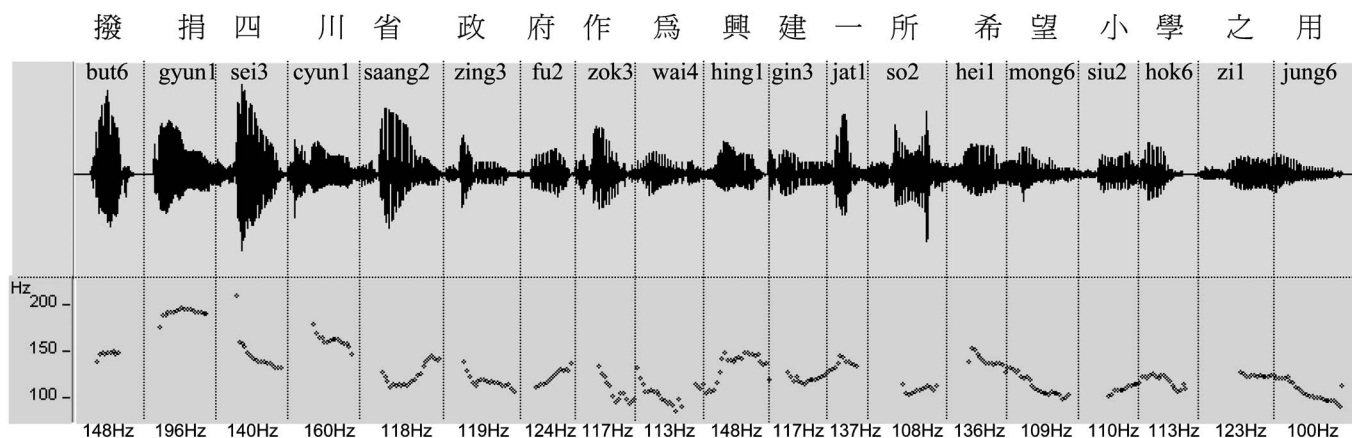
FIG. 3. The F0 contours of a continuous Cantonese sentence spoken by a male speaker. The upper panel shows the waveform and the lower panel gives the corresponding F0 plot. The content of the sentence is given in both Chinese characters and their Jyut Ping transcriptions. The numbers at the bottom are the average F0 values as computed from the middle section of the syllable-wide F0 contours.

Qian *et al.*: Tone recognition using supratone models

(a)

(b)
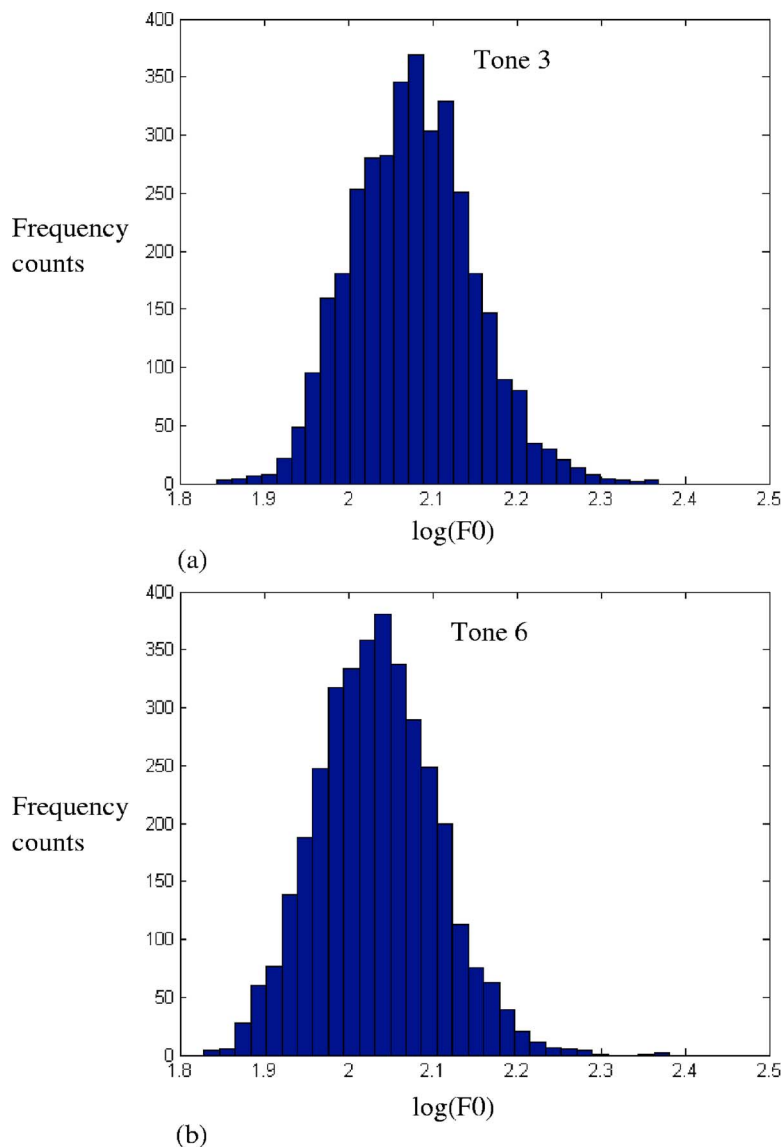
FIG. 5. (Color online) Histograms of the F0 levels of (a) Tone 3 and (b) Tone 6. The data are obtained from all of the syllable pairs with the combination of Tone 3 followed by Tone 6 in the training utterances of the 34 male speakers in CUSENT. The log(F0) value is computed from the middle section of the syllable's F0 contour.

C. Importance of relative pitch difference

The six tones of Cantonese can be roughly categorized as level tones or rising tones, according to the shapes of tone contours. This is unlike Mandarin, in which all four basic tones have distinctive contour shapes, namely, high-level, mid-rising, falling-rising, and high-falling (Xu, 1997). Discrimination between the Cantonese tones relies more on the heights than on the shapes of the pitch contours. Bauer and Benedict (1997) pointed out that the height of a tone is not an absolute feature. It is the relative pitch difference, rather than the absolute F0 level, which makes the tones identifiable and distinguishable.

Given the large variation of F0, it is questionable whether tones can retain their distinctive pitch contours in natural speech. Let us pick an arbitrary pair of syllables and compare their pitch heights. If the syllables are not in neighboring positions, their relative pitch heights may not follow the order as in Figs. 1 and 2. In the example of Fig. 3, the third syllable, which carries Tone 3, has a noticeably higher F0 than the 18th syllable, which carries Tone 1. However,

from the same example it can be seen that the relative pitch heights are mostly preserved for neighboring tones. It is very unlikely, if not impossible, that a Tone 4 syllable would have a higher F0 than a Tone 3 syllable in its near neighborhood.

Using the same set of utterances as in Fig. 4, all of the syllable pairs with the combination of Tone 3 followed by Tone 6 are analyzed. The pitch level of each syllable is represented by the average logarithmic F0 value of its middle section. The statistical distributions of the absolute F0 levels of Tone 3 syllables and Tone 6 syllables are plotted as in Figs. 5(a) and 5(b), respectively. The statistical mean for Tone 3 is slightly higher than that for Tone 6, and the F0 ranges of the two tones are so wide that they greatly overlap each other. In other words, it is difficult to distinguish them based on the absolute F0 level. Figure 6 gives the distribution of the log(F0) differences of the tone pairs, which are computed by subtracting the log(F0) value of the succeeding syllable (Tone 6) from that of the preceding syllable (Tone 3). A positive difference value means that the preceding syllable has a higher F0 value than the succeeding syllable. The
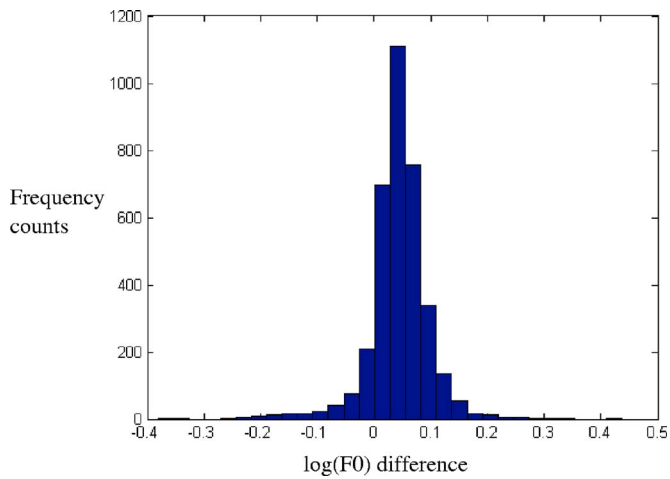
FIG. 6. (Color online) Histogram of the log(F0) differences between Tone 3 and Tone 6. The data are obtained from all of the syllable pairs with the combination of Tone 3 followed by Tone 6 in the training utterances of the 34 male speakers in CUSENT. The F0 difference is computed by subtracting the logarithmic F0 level of the succeeding syllable (Tone 6) from that of the preceding syllable (Tone 3).

FIG. 7. The process of extracting feature vectors for ditone modeling.

distribution shows that the data have positive values in most cases. It is clear that the relative F0 levels of Tone 3 and Tone 6 are well retained.

To effectively recognize a tone in continuous Cantonese speech, it is not only the F0 contour of the syllable that is useful. Those of its neighboring syllables play an important role in serving as a reference for the pitch level. This motivates us to extend the window for tone modeling beyond the scope of a single syllable. On the other hand, the control of F0 in producing individual tones is confounded by the control of F0 in producing intonation. The intonational effect can be alleviated by F0 normalization, which requires the consideration of long stretches of speech. If the relative F0 levels between members of abutting tone pairs are kept largely distinct, it may be possible to improve tone recognition by modeling the tone contours in succession, rather than by normalization with respect to their locations within an intonational phrase. The proposed approach of supratone modeling is explained in detail in Sec. IV.

IV. SUPRATONE MODELING FOR CANTONESE

A. Definition of supratone units

A supratone unit refers to the concatenation of multiple tones in succession. In other words, it covers more than one syllable. In this study, we focus on ditone and tritone units, i.e., the concatenation of two and three neighboring tones, respectively. By including multiple tone contours in a single modeling unit, the relative heights of the tones can be explicitly captured. As discussed in Sec. III C, this information is especially useful for Cantonese. As there are six basic tones in Cantonese, the total number of ditone and tritone units is 36 and 216, respectively.

The notion of supratone modeling is fundamentally different from that of conventional context-dependent modeling of monotone units (Wang *et al.*, 1997; Lee *et al.*, 2002a). A context-dependent monotone model is defined based on the
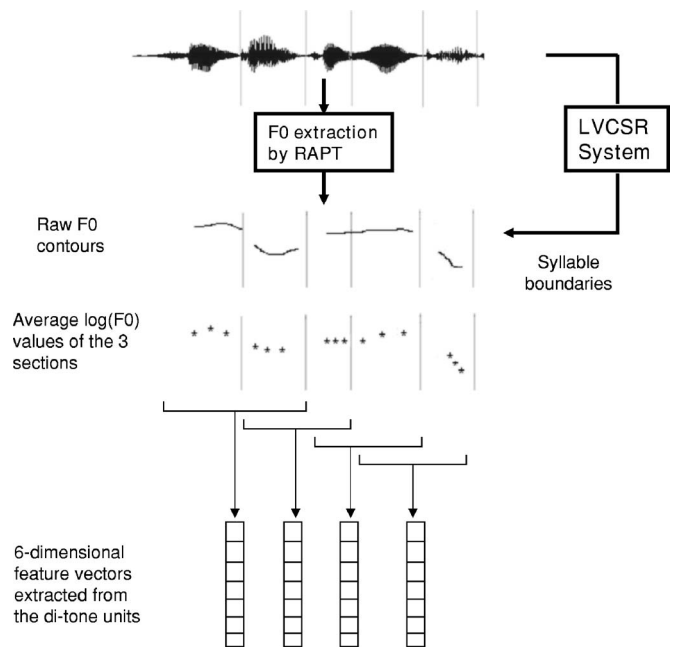
phonetic context, rather than the acoustic context. The term "context" refers to the categorical identities of neighboring tones, and the exact acoustic realization of these tones is not considered. A supratone model captures the tonal contextual effect by using not only the tone identities, but also the acoustic features, i.e., F0 contours, of the neighboring syllables. In this way, the tonal context can be characterized more precisely than by simply relying on the linguistically defined tone categories. Similar ideas were adopted in the previous study by Zhang *et al.* (2004).

B. Extraction of feature parameters for supratone modeling

Given a continuous speech utterance, a sequence of F0 values are computed on a short-time basis, in this case every 0.01 s, using the RAPT (Talkin, 1995). The F0 contour covers only the final segment of a syllable. To align the extracted F0 values with individual syllables, a Cantonese LVCSR system is used. This system uses initials and finals as the fundamental units for acoustic modeling (Wong *et al.*, 1999). With these acoustic models, speech segmentation at initial-final level is generated as a by-product of the recognition process (Lee *et al.*, 2002a). For training data, since the content of the utterance is given, a more accurate segmentation can be obtained by HMM forced alignment (Young *et al.*, 2001).

Each tone contour is divided evenly into three sections, and the average logarithmic F0 value of each section is computed. Thus, each tone is represented by only three coarsely sampled points. In ditone and tritone modeling, the feature vector is composed of six and nine components, respectively. Figure 7 explains the feature extraction process for ditone modeling.

Previous research has shown that the height and shape of F0 contours, rather than the exact values at densely sampled points, are critical for the recognition of Chinese tones. For

Mandarin tone recognition, the feature parameters that are used to represent a tone typically include three to five F0 values that span the syllable, or a few parameters that describe the geometrical shape and overall height of the contour (Chen and Wang, 1995; Lin and Lee, 2003; Zhang *et al.*, 2005). In Tian *et al.* (2004), it was suggested that a rough sketch of the F0 contour is more appropriate for Mandarin tone recognition than a fine description. From a statistical modeling point of view, it is also desirable to have a relatively small number of feature parameters.

C. Overlapping of supratone units

As shown in Fig. 7, each pair of neighboring tones forms a ditone unit. To cover all tonal transitions in the utterance, the ditone units are overlapped. There are a number of advantages of such overlapping. First, for the same utterance, the number of overlapping ditone units is approximately double of what it would be without overlapping. This is desirable for the statistical modeling of tones, especially when the amount of training data is limited. Second, the overlap provides an additional constraint to the tone recognition process. For example, given that there are only four distinct level tones in Cantonese, it is not permitted to have more than three successive ditone units with an upward change of pitch level. Details of the proposed tone recognition algorithm are given in Sec. IV E.

D. Statistical modeling using Gaussian mixture models

Various techniques of pattern recognition have been applied to the classification of tones, such as support vector machines (Peng and Wang, 2005), artificial neural networks (Lee *et al.*, 1995; Chen and Wang, 1995), and hidden Markov models (Wang *et al.*, 1997). In this study, the classification of supratone feature vectors is accomplished with a set of Gaussian mixture models (GMMs). A GMM approximates a multimodal distribution by a weighted combination of a finite number of unimodal Gaussian distributions (Huang *et al.*, 2001). It can be regarded as a single-state HMM. It provides a probabilistic output that can be readily integrated with the state-of-the-art HMM-based ASR framework (Qian *et al.*, 2004b).

Each supratone unit is modeled by a dedicated GMM. The model parameters, which are the mixture weights, mean vectors, and covariance matrices, are estimated from a large number of training feature vectors extracted from speech segments that carry this supratone unit. Diagonal covariance matrices are assumed so that the computational complexity is practically affordable. The *expectation-maximization* (EM) algorithm is used for training of the GMMs (McLachlan and Krishnan, 1997). The mean vectors and covariance matrices are initialized by a single-iteration k-means clustering algorithm using all training data and equal mixture weights. The number of mixture components in each GMM is fixed at the beginning. With multiple iterations of the EM algorithm, the number of components is gradually increased until the output probability does not increase noticeably or the number of iterations reaches a preset limit.

TABLE I. Mean vectors of ditone GMM trained with male speech. The ditone unit is the combination of Tone 1 followed by Tone 4. Each row shows one of the mean vectors. Each mean vector contains six F0 values.

Mixture component	Elements in each trained mean vector (Hz)					
1	144	145	144	113	102	98
2	126	131	132	112	108	115
3	147	146	147	115	102	96
4	129	143	142	113	108	103

Each Gaussian distribution in a GMM can be considered to represent a specific subcluster of the feature space. Table I lists the estimated mean vectors of a ditone GMM that represents the combination of Tone 1 followed by Tone 4. The training data are from 34 male speakers of the CUSENT database (see Sec. V A). In this example, the GMM is forced to have exactly four mixture components. Each mean vector consists of six logarithmic F0 values that represent different sections of the F0 contours. The first three elements in the vector represent the F0 contours of Tone 1 and the next three elements represent those of Tone 4. For ease of interpretation, the logarithmic F0 values are converted to linear values (in hertz). It can be observed that the trained Gaussian distributions reflect the F0 difference between the two tones. There are also noticeable differences among the individual mean vectors, which correspond to different contextual variations of the ditone contours.

E. Tone recognition based on supratone models

The following description of the proposed tone recognition algorithm is based on ditone models. The algorithm with tritone models is very similar, but is more complicated. Let i-j denote a ditone model that is trained for the combination of Tone i followed by Tone j, where $i,j=1,2,\ldots,6$. There are a total of 36 ditone models. In addition, the tones at the beginning of utterances are specially modeled by another six GMMs, which are denoted as sil-j, where $j=1,2,\ldots,6$. The process of tone recognition is formulated as a problem of searching for the most likely sequence of ditone models given a sequence of feature vectors derived from the overlapped ditone F0 contours. An example is shown in Fig. 8. In this example, the input utterance contains four syllables, from which four consecutive feature vectors are obtained. The ditone models form a search space. The transition between a pair of ditone units is allowed only if they overlap with the same tone. This is illustrated by a dashed line in Fig. 8. Any path that passes through admissible transitions from the rightmost column to the leftmost column of the model space represents a valid candidate for the tone sequence, as illustrated by the thick solid line in Fig. 8. Of all of the candidate paths, the one with the highest probability is selected as the result of tone recognition. Each ditone model produces an output probability for each feature vector. The overall probability of a valid path is equal to the accumulated probability of all of the models on the path. For this task, an efficient search can be accomplished using standard dynamic programming or the Viterbi search algorithm (Ney and Ortmanns, 1999). This search is a forward search that starts
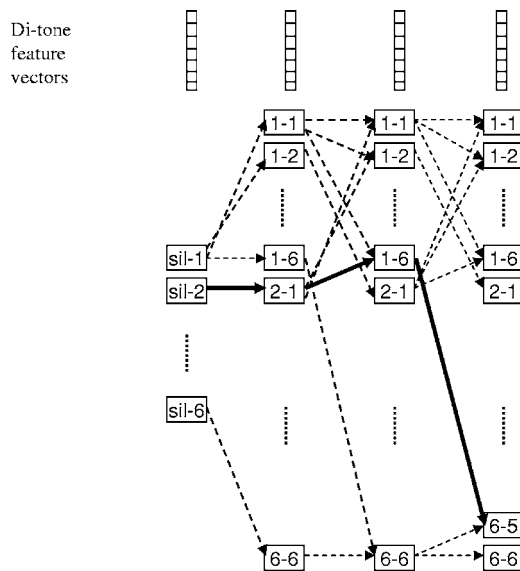
FIG. 8. Tone recognition as a process of searching for the most likely ditone sequence. Each box represents a ditone model. The dashed lines are the admissible transitions. The thick solid line shows a valid path, which corresponds to the tone sequence: 2 1 6 5.

from the first input feature vector. At each time step, a new feature vector is involved and a partially optimal path is determined. The search terminates when the last feature vector is processed.

In Cantonese, there are about 600 base syllables, while the total number of tonal syllables is about 1800. In other words, more than half of the syllable-tone combinations are not allowed. The proposed algorithm for tone recognition takes advantage of such phonological constraints to produce a linguistically meaningful output. While the Cantonese LVCSR system is utilized to provide syllable boundaries for the extraction of tone features, it also produces a hypothesized sequence of base syllables. In the process of searching for the best tone sequence, if the combination of a tone and a hypothesized base syllable is not allowed, this tone will have no chance of being included in the optimal path.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Speech database

Speaker-independent tone recognition experiments are carried out using CUSENT, which is part of a series of Cantonese speech databases developed at the Chinese University of Hong Kong (CUCorpora, 2001; Lee *et al.*, 2002b). CUSENT is a read speech database. The content of the corpus was selected from five local newspapers. The training data consists of 20 000 phonetically rich continuous sentences that were uttered by 34 male and 34 female speakers. The test data comprises 1200 sentences from a different set of 12 speakers (6 male and 6 female). All of the training and test utterances were transcribed manually in Jyut Ping. Table II shows the distributions of the six tones in the test utterances. Tone 1 occurs the most frequently and Tone 5 the least frequently. The average length of utterances in CUSENT is about 10 syllables.

TABLE II. Distribution of the six tones in the test utterances in CUSENT.

Tone	Distribution (%)
1	22.2
2	13.3
3	17.1
4	18.6
5	7.1
6	21.7

B. Tone recognition with ditone models

The total number of ditone units is 42, including the 36 tone pairs and 6 utterance-initial tones. The ditone GMMs are built based on the 20 000 training utterances in CUSENT. The number of mixture components in the ditone GMMs varies from 8 to 38, depending on the amount of training data. Tone recognition is carried out for each test utterance as detailed in Sec. IV E.

In the proposed recognition algorithm, a Cantonese LVCSR system is needed to provide the syllable boundaries and syllable hypotheses. This system was described in detail in Wong *et al.* (1999) and Choi *et al.* (2000). With the syllable boundaries produced by the LVCSR system, the overall accuracy of tone recognition using the ditone models is 74.68%. If we assume perfect syllable recognition and use the syllable segmentation produced by HMM forced alignment, the recognition accuracy increases to 79.18%. This can be considered as the best performance that the proposed method can attain.

Table III shows the confusion patterns among the six tones as attained in the case of perfect syllable recognition. Similar patterns are observed when the syllable hypotheses are obtained by unconstrained LVCSR. Among the six tones, the highest and the lowest accuracy, 96.40% and 58.94%, are attained for Tone 1 and Tone 5, respectively.

Fok-Chan (1974) reported a perceptual study of the six Cantonese tones. The speech stimuli were syllables uttered in isolation by two speakers. The overall accuracy of recognition (over 511 subjects) was about 90%. In Ma *et al.* (2005), the effect of tone context on Cantonese tone perception was studied. The recognition accuracy varied from 78% to 98%, depending on the extent of contextual information that was provided. The tone recognition accuracy attained by our pro-

TABLE III. Confusion matrix and overall accuracy of tone recognition by ditone models. Perfect syllable recognition is assumed and syllable segmentation by HMM forced alignment is used.

		Recognized tone						Accuracy	
		1	2	3	4	5	6	Individual tones (%)	Overall
Correct tone	1	2061	6	35	4	3	29	96.40	79.18%
	2	49	1008	66	33	47	76	78.81	
	3	96	64	1192	17	56	215	72.68	
	4	14	43	57	1399	51	227	78.11	
	5	7	117	21	44	399	89	58.94	
	6	54	69	232	101	79	1552	74.37	

Qian *et al.*: Tone recognition using supratone models

TABLE IV. The ten most frequently occurring ditone errors. "1-3→1-6" means that the unit "1-3" is recognized as "1-6."

10 most frequent ditone recognition errors
1-3 → 1-6
6-1 → 3-1
1-6 → 1-3
6-4 → 3-4
1-4 → 1-6
4-1 → 6-1
3-1 → 6-1
3-4 → 6-4
6-6 → 6-3
4-6 → 4-3

TABLE V. Comparison of tone recognition performance of GMM based context-dependent monotone models and the proposed supratone models.

Approach	No. of feature parameters per syllable	No. of models required	No. of mixture components in the GMMs	Tone recognition accuracy (%)
Context-dependent monotone modeling	3	42	13–64	70.05
Ditone modeling	3	42	8–38	74.68
Tritone modeling	3	258	1–13	75.59

posed method is considered to be fairly high, given that the training and test data are continuous utterances spoken by a large number of different speakers.

Table III shows that the most confusable pair of tones are Tone 3 and Tone 6. As we can see in Fig. 2, the F0 difference between the two tones is about 20 Hz even for isolated syllables uttered by a single speaker. In continuous sentences that are spoken by many different speakers, the F0 ranges of these two tones overlap greatly, as shown in Figs. 4 and 5. The confusion patterns found by the perceptual studies of Fok-Chan (1974) and Ma *et al.* (2005) are very similar to our observations.

The performance of the proposed method can also be assessed in terms of the percentage of correctly recognized ditone units. The ditone accuracy is 68.29% in the case of perfect syllable recognition. Table IV lists the 10 most frequent ditone errors. These errors occur mostly between ditone units that have the same direction of pitch movement, e.g., a high-pitch tone followed by a low-pitch tone. We find very few cases of confusion between ditone units that have opposite directions of pitch movement. For example, there is only one case of confusion between the units 6-3 and 3-6, although Tone 3 and Tone 6 are highly confusable individually.

C. Tone recognition with tritone models

The total number of units for tritone modeling is 258. The tritone GMMs are trained in the same way as the ditone GMMs. The number of mixture components in the tritone GMMs ranges between 1 and 13. The overall accuracy of tone recognition is 75.59%, which is slightly better than that of using the ditone models. A tritone unit covers a longer time span than a ditone unit. Each tritone unit overlaps its immediately neighboring unit by two tones, and there is an overlap between three consecutive tritone units. This imposes more stringent constraints on the search for the optimal tone sequence than when the ditone models are used. The experimental results show that tritone modeling is only marginally superior to ditone modeling. Given the same speech database, the number of training tokens that is needed for the tritone units is about one-sixth that for the ditone units. This may affect the effectiveness of the statistical modeling, and thus limit the performance of the tritone models.

D. Comparison with context-dependent monotone models

Context-dependent monotone modeling has been a commonly used approach to automatic tone recognition in continuous speech. In our experiment, we use 42 GMMs, each of which represents a right-context-dependent tone. The feature vector extracted from each tone contains three logarithmic F0 values. Table V compares the supratone models with that of the GMM based context-dependent monotone models, in terms of the tone recognition accuracy, the dimension of feature vectors, the number of models required, and the number of mixture components in the GMMs. Figure 9 compares the recognition accuracy on individual tones between the ditone models and the monotone models. The accuracy for Tones 3 and 6 are improved significantly from 60.78% and 63.47% to 69.9% and 71.9%, respectively. This clearly shows the advantage of supratone models in dealing with the confusing level tones of Cantonese.

E. Issues on the selection of feature parameters

Based on the framework of supratone models, we have made an initial attempt to use duration features for Cantonese tone recognition. The duration of a tone is represented by the duration of the final part of the syllable. The tone duration is normalized with respect to the average tone duration over the utterance. By including the normalized tone duration as additional components of the supratone feature vector, the tone recognition accuracy with ditone models is improved from 74.68% to 75.28%. These results are preliminary and not considered to be conclusive. The duration of a tone is determined not only by the tone identity, but also by other factors, such as the phonetic properties, speaking rate, and part of speech. Further investigation is needed to devise more discriminative duration features for tone identification.
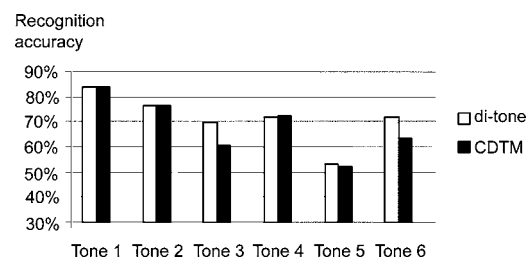
FIG. 9. Performance comparison between ditone models and context-dependent monotone models.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Qian *et al.*: Tone recognition using supratone models   2943

In Peng and Wang (2005), the tone feature vector extracted for each syllable consists of 20 components, which include the normalized F0, the temporal slope of F0, energy, and duration. An accuracy of 71.5% was attained for CUSENT. For Cantonese tone recognition, proper normalization of F0 is important to reduce tone-irrelevant F0 variation, e.g., intonation and pitch change of speakers. F0 normalization can be done with a long-term or even an utterance-wide mean F0 value (Lee *et al.*, 2002a; Peng and Wang, 2005). The proposed ditone models can attain an accuracy of 74.68% using only six feature parameters for each syllable, and use the extracted F0 values directly without requiring any normalization procedure. We have tried to apply various F0 normalization techniques to the logarithmic F0 feature parameters for ditone modeling. No noticeable improvement could be attained. It seems that by explicitly characterizing the relative pitch level of neighboring tones, the confusion caused by long-term F0 variation is at least partially resolved.

F. Use of tone information in LVCSR

It is believed that the accuracy of Cantonese LVCSR can be improved by using correct tone information. However, perfect tone recognition is not possible. For speaker-independent tone recognition in continuous Cantonese speech, the attainable recognition accuracy is typically 70%–80%, as reported in this study and the work by others (Peng and Wang, 2005). With this performance level, the output of tone recognition can be of use for Cantonese LVCSR, as demonstrated in Lee *et al.* (2002a) and Peng and Wang (2005). In our recent studies (Qian *et al.*, 2004a, b), a number of different approaches have been developed to integrate the probabilistic output of supratone models into the search process of Cantonese LVCSR. Tone information is used to provide a confidence measure that can be used either directly during the search or for the rescoring of character hypotheses. As shown in Table III, there are large differences in the recognition accuracy for different tones. In other words, the recognition results for some of the tones are more reliable than those of the others. Our approach is to keep the character hypotheses that are deemed reliable intact and exploit the reliable tone information to disambiguate confusing character hypotheses. This leads to a noticeable improvement in the overall performance of LVCSR (Qian *et al.*, 2004b).

VI. CONCLUDING REMARKS

Cantonese tone recognition has long been considered a difficult task because the number of tones is relatively large and the pitch contours of different tones are confusable. Nevertheless, the pitch contrast among tones in near neighborhood remains an effective feature for tone recognition. The design of supratone models is intended to capture such pitch differences explicitly. The results of tone recognition experiments show that the supratone models are more effective in resolving the confusion between the level tones of Cantonese than the conventional context-dependent monotone models. Moreover, the proposed method requires a relatively small number of feature parameters to represent each tone, and perform better than other methods reported previously.

The approach of supratone modeling in its existing implementation can be improved in many aspects. The accuracy of tone recognition is at best about 79%. Further investigations are needed to more effectively reduce the tone-irrelevant contextual variation of the F0, and to exploit other discriminative acoustic features, including duration and energy.

[1]Here the International Phonetic Alphabet is used to make the pronunciation comprehensible to general readers. In subsequent sections, Cantonese syllables and tones are transcribed in Jyut Ping, which is a Romanization system devised by the Linguistic Society of Hong Kong (LSHK, 1997).

[2]The choice of nine or ten tones depends on whether the high level and high falling tones are treated as distinct. Cheung (1986) and Bauer and Benedict (1997) found that high falling tones are most commonly pronounced as high level tones in Hong Kong Cantonese. Thus in this study it is assumed that our speech materials do not contain any high falling tones.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology*, Trends in Linguistics: Studies and Monographs, Vol. 102, (Mouton de Gruyer, Berlin, NY).

Blicher, D. L., Diehl, R., and Cohen, L. B. (**1990**). "Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement," J. Phonetics **18**, 37–49.

Chang, P. C., Sun, S. W., and Chen, S. H. (**1990**). "Mandarin tone recognition by multi-layer perception," Proceedings ICASSP, Albuquerque, USA, Vol. 1, pp. 517–520.

Chao, Y. R. (**1947**). *Cantonese Primer* (Harvard University Press, Cambridge).

Chen, M. Y. (**2000**). *Tone Sandhi: Patterns across Chinese Dialects* (Cambridge University Press, Cambridge).

Chen, S. H., and Wang, Y. R. (**1995**). "Tone recognition of continuous Mandarin speech based on neural networks," IEEE Trans. Speech Audio Process. **3**, 146–150.

Cheung, K. H. (**1986**). "The Phonology of Present-Day Cantonese," Ph.D. dissertation, University College, London.

Ching, P. C., Lee, T., and Zee, E. (**1994**). "From phonology and acoustic properties to automatic recognition of Cantonese," Proceedings of International Symposium on Speech, Image Processing and Neural Networks, Hong Kong, Vol. **1**, pp. 127–132.

Choi, W. N., Wong, Y. W., Lee, T., and Ching, P. C. (**2000**). "Lexical-tree decoding with a class-based language model for Chinese speech recognition," Proceedings ICSLP, Beijing, pp. 174–177.

CUCorpora: Cantonese Spoken Language Resources (**2001**). http://dsp.ee.cuhk.edu.hl/speech/ (last viewed on 3 January 2007).

Cutler, A., and Chen, H.-C. (**1997**). "Lexical tone in Cantonese spoken-word processing," Percept. Psychophys. **59**, 165–179.

Fok-Chan, Y. Y. (**1974**). *A Perceptual Study of Tones in Cantonese* (University of Hong Kong Press, Hong Kong).

Hashimoto, O. K. Y. (**1972**). *Studies in Yue Dialects 1: Phonology of Cantonese* (Cambridge University Press, New York).

Hon, H. W., Yuan, B., Chow, Y. L., Narayan, S., and Lee, K. F. (**1994**). "Towards large vocabulary Mandarin Chinese speech recognition," Proceedings ICASSP, Adelaide, Australia, Vol. 1, pp. 545–548.

Huang, X., Acero, A., and Hon, H. W. (**2001**). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development* (Prentice Hall, Englewood Cliffs, NJ).

Lee, T., Ching, P. C., Chan, L. W., Mak, B., and Cheng, Y. H. (**1995**). "Tone recognition of isolated Cantonese syllables," IEEE Trans. Speech Audio Process. **3**, 204–209.

2944   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Qian *et al.*: Tone recognition using supratone models

Lee, T., Kochanski, G., Shih, C., and Li, Y. J. (**2002c**). "Modeling tones in continuous Cantonese speech," in Proceedings ICSLP, Denver, USA, pp. 2401–2404.

Lee, T., Lau, W., Wong, Y. W., and Ching, P. C. (**2002a**). "Using tone information in Cantonese continuous speech recognition," ACM Trans. on Asian Language Information Processing **1**, 83–102.

Lee, T., Lo, W. K., Ching, P. C., and Meng, H. (**2002b**). "Spoken language resources for Cantonese speech processing," Speech Commun. **36**, 327–342.

Li, Y. J., Lee, T., and Qian, Y. (**2004**). "Analysis and modeling of F0 contours for Cantonese text-to-speech," ACM Trans. Asian Language Information Processing **3**, 169–180.

Lin, C. H., Wu, C. H., Ting, P. Y., and Wang, H. M. (**1996**). "Frameworks for recognition of Mandarin syllables with tones using sub-syllabic units," Speech Commun. **18**, 175–190.

Lin, W. Y., and Lee, L. S. (**2003**). "Improved tone recognition for fluent Mandarin speech based on new inter-syllabic features and robust pitch extraction," Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding, Virgin Islands, USA, pp. 237–242.

Linguistic Society of Hong Kong (LSHK) (**1997**). *Hong Kong Jyut Ping Characters Table* (Linguistic Society of Hong Kong Press, Hong Kong).

Liu, J. (**2001**). "Tonal behavior in some tone languages," Ph.D dissertation, City University of Hong Kong, Hong Kong.

Ma, K. Y., Ciocca, V., and Whitehill, T. (**2005**). "Contextual effect on perception of lexical tones in Cantonese," Proceedings ICSLP, Lisbon, Portugal, pp. 401–404.

McLachlan, G., and Krishnan, T. (**1997**). *The EM Algorithm and Extensions* (Wiley, New York).

Ney, H., and Ortmanns, S. (**1999**). "Dynamic programming search for continuous speech recognition," IEEE Signal Process. Mag. **16**, 64–83.

Peng, G., and Wang, W. S. Y. (**2005**). "Tone recognition of continuous Cantonese speech based on support vector machines," Speech Commun. **45**, 49–62.

Pike, K. L. (**1948**). *Tone Languages* (University of Michigan Press, Ann Arbor).

Qian, Y., Lee, T., and Soong, F. K. (**2004a**). "Use of tone information in continuous Cantonese speech recognition," Proceedings of Speech Prosody, Nara, Japan, pp. 587–590.

Qian, Y., Lee, T., and Soong, F. K. (**2004b**). "Tone information as a confidence measure for improving Cantonese LVCSR," Proceedings ICSLP, Jeju, Korea, Vol. 3, pp. 1965–1969.

Talkin, A. D. (**1995**). "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, edited by W. B. Kleijn and K. K. Paliwal (Elsevier Science, Amsterdam).

Tian, Y., Zhou, J. L., Chu, M., and Chang, E. (**2004**). "Tone recognition with fractionized models and outlined features," Proceedings ICASSP, Montreal, Canada, Vol. 1, pp. 105–108.

Wang, H. M., Ho, T. H., Yang, R. C., Shen, J. L., Bai, B. R., Hong, J. C., Chen, W. P., Yu, T. L., and Lee, L.-S. (**1997**). "Complete recognition of continuous Mandarin speech for Chinese language with very large vocabulary using limited training data," IEEE Trans. Speech Audio Process. **5**, 195–200.

Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.

Wong, Y. W., Chow, K. F., Lau, W., Lo, W. K., Lee, T., and Ching, P. C. (**1999**). "Acoustic modeling and language modeling for Cantonese LVCSR," Proceedings Eurospeech, Budapest, Hungary, pp. 1091–1094.

Xu, Y. (**1997**). "Contextual tonal variation on Mandarin," J. Phonetics **25**, 61–83.

Xu, Y. (**2001**). "Sources of tonal variations in connected speech," Journal of Chinese Linguistics Monograph Series **17**, 1–31.

Yang, W. J., Lee, J. C., Chang, Y. C., and Wang, H. C. (**1988**). "Hidden Markov model for Mandarin lexical tone recognition," IEEE Trans. ASSP **36**, 988–992.

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., Valtchev, V., and Woodland, P. (**2001**). *The HTK Book* (for HTK Version 3.1) (Cambridge University, Cambridge, UK).

Yuan, J. (**1960**). *Hanyu Fangyan Gaiyao [An Introduction to Chinese Dialectology]* (Wenzi Gaige Chubanshe, Beijing).

Zhang, J. S., Nakamura, S., and Hirose, K. (**2004**). "Tonal contextual F0 variations and anchoring based discrimination," Proceedings of Speech Prosody, Nara, Japan, pp. 525–528.

Zhang, J. S., Nakamura, S., and Hirose, K. (**2005**). "Tone nucleus-based multi-level robust acoustic tonal modeling of sentential F0 variations for Chinese continuous speech tone recognition," Speech Commun. **46**, 440–454.

Statistical voice activity detection based on integrated bispectrum likelihood ratio tests for robust speech recognition

J. Ramírez,[a] J. M. Górriz, and J. C. Segura
Deptament of Signal Theory, Networking and Communications, University of Granada, Granada, Spain

Currently, there are technology barriers inhibiting speech processing systems that work in extremely noisy conditions from meeting the demands of modern applications. These systems often require a noise reduction system working in combination with a precise voice activity detector (VAD). This paper shows statistical likelihood ratio tests formulated in terms of the integrated bispectrum of the noisy signal. The integrated bispectrum is defined as a cross spectrum between the signal and its square, and therefore a function of a single frequency variable. It inherits the ability of higher order statistics to detect signals in noise with many other additional advantages: (i) Its computation as a cross spectrum leads to significant computational savings, and (ii) the variance of the estimator is of the same order as that of the power spectrum estimator. The proposed approach incorporates contextual information to the decision rule, a strategy that has reported significant benefits for robust speech recognition applications. The proposed VAD is compared to the G.729, adaptive multirate, and advanced front-end standards as well as recently reported algorithms showing a sustained advantage in speech/nonspeech detection accuracy and speech recognition performance.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2714915]

I. INTRODUCTION

The emerging applications of speech technologies (particularly in mobile communications, robust speech recognition, or digital hearing aid devices) often require a noise reduction scheme working in combination with a precise voice activity detector (VAD).[1] During the last decade numerous researchers have studied different strategies for detecting speech in noise and the influence of the VAD decision on speech processing systems.[2–8] This task can be identified as a statistical hypothesis testing problem and its purpose is the determination to which category or class a given signal belongs. The decision is made based on an observation vector, frequently called feature vector, which serves as the input to a decision rule that assigns a sample vector to one of the given classes. The classification task is often not as trivial as it appears since the increasing level of background noise degrades the classifier effectiveness and causes numerous detection errors.[9,10]

The nonspeech detection algorithm is an important and sensitive part of most of the existing single-microphone noise reduction schemes. Well-known noise suppression algorithms[11,12] such as Wiener filtering (WF) or spectral subtraction, are widely used for robust speech recognition being the VAD critical in attaining a high level of performance. These techniques estimate the noise spectrum during nonspeech periods in order to compensate for the harmful effect of the noise on the speech signal. The VAD is even more critical for nonstationary noise environments since the statistics of the background noise must be updated. An example of

such a system is the ETSI standard for distributed speech recognition that incorporates noise suppression methods. The so-called advanced front-end (AFE)[13] considers an energy-based VAD in order to estimate the noise spectrum for Wiener filtering and a different VAD for nonspeech frame dropping (FD).

On the other hand, a VAD achieves silence compression in modern mobile telecommunication systems reducing the average bit rate by using the discontinuous transmission mode. Many practical applications, such as the Global System for Mobile Communications (GSM) telephony, use silence detection and comfort noise injection for higher coding efficiency. The International Telecommunication Union (ITU) adopted a toll-quality speech coding algorithm known as G.729 to work in combination with a VAD module in DTX mode. The recommendation G.729 Annex B[4] uses a feature vector consisting of the linear prediction spectrum, the full-band energy, the low-band (0–1 kHz) energy, and the zero-crossing rate. Another standard for DTX is the ETSI (adaptive multirate) AMR speech coder[3] developed by the Special Mobile Group for the GSM system. The AMR standard specifies two options for the VAD to be used within the digital cellular telecommunications system. In option 1, the signal is passed through a filterbank and the subband energies are calculated. The VAD decision depends on a measure of the signal-to-noise ratio, the output of a pitch detector, a tone detector, and the correlated complex signal analysis module. An enhanced version of the original VAD is the AMR option 2 that uses parameters of the speech encoder that are more robust to the environmental noise than AMR1 and G.729. These VADs have been used extensively in the

a)Electronic mail: javierrp@ugr.es

open literature as a reference for assessing the performance of new algorithms.

Most of the algorithms for detecting the presence of speech in a noisy signal only exploit the power spectral content of the signals and require knowledge of the noise power spectral density.[6,8,14,15] One of the most important disadvantages of these approaches is that no *a priori* information about the statistical properties of the signals is used. Higher order statistics methods rely on an *a priori* knowledge of the input processes and have been considered for VAD since they can distinguish between Gaussian signals (which have a vanishing bispectrum) from non-Gaussian signals. However, the main limitations of bispectrum-based techniques are that they are computationally expensive and the variance of the bispectrum estimators is much higher than that of power spectral estimators for identical data record size. These problems were addressed by Tugnait,[16,17] who showed a computationally efficient and reduced variance statistical test based on the integrated polyspectra for detecting a stationary, non-Gaussian signal in Gaussian noise. This paper shows an effective VAD based on a likelihood ratio test (LRT) defined on the integrated bispectrum of the noisy speech. The proposed approach also incorporates contextual information to the decision rule, a strategy first proposed in Ref. 18 that has reported significant benefits for different applications including robust speech recognition.[19–22] The paper includes a careful derivation of the LRT previously addressed in Refs. 23 and 24, an alternative approach based on block partitioning and averaging of the integrated bispectra, and its efficient implementation using the contextual LRT. The paper is organized as follows. Section II reviews the definition and fundamental properties of third-order cumulants and bispectrum. Section III suggests the use of integrated bispectrum for a reduced variance estimation while maintaining the benefits of higher order statistics for detection. Section IV shows the definition of the VAD based on contextual integrated bispectrum LRTs. Section V shows two different methods for integrated bispectrum estimation and LRT definition and their analysis and comparison is shown in Sec. VI. Section VII shows the receiver operating characteristic (ROC) curves and speech recognition experiments that are used to evaluate the proposed method and to compare its performance to ITU-T G.729, ETSI AMR and AFE, as well as to other recently reported VADs. Finally, Sec. VIII summarizes the conclusions of this work.

II. BISPECTRUM

The bispectrum of a deterministic, continuous-time signal $x(t)$ is defined as[25,26]

$$B(\omega_1, \omega_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C_{3x}(\tau_1, \tau_2)$$

$$\times \exp\{-j(\omega_1 \tau_1 + \omega_2 \tau_2)\} d\tau_1 d\tau_2, \quad (1)$$

where

$$C_{3x}(\tau_1, \tau_2) = E\{x^*(t)x(t + \tau_1)x(t + \tau_2)\}$$

$$= \int_{-\infty}^{+\infty} x^*(t)x(t + \tau_1)x(t + \tau_2)dt \quad (2)$$

is the third-order cumulant of $x(t)$, and $\omega = 2\pi f$ with normalized frequency f. By the symmetry properties, the bispectrum of a real signal is uniquely defined by its values in the triangular region $0 \leq \omega_2 \leq \omega_1 \leq \omega_1 + \omega_2 \leq \pi$, provided that there is no bispectral alias. In a similar fashion, the bispectrum of a discrete-time signal is defined as the two-dimensional (2D) Fourier transform:

$$B_x(\omega_1, \omega_2) = \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{3x}(i,k)\exp\{-j(\omega_1 i + \omega_2 k)\}. \quad (3)$$

Note that from the above-presented definition the third-order cumulant can be expressed as

$$C_{3x}(i,k) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} B_x(\omega_1, \omega_2)$$

$$\times \exp\{j(\omega_1 i + \omega_2 k)\} d\omega_1 d\omega_2 \quad (4)$$

using the inverse Fourier transform.

Figure 1 shows the differences between cumulants and bispectrum function (magnitude and phase) for a frame containing noise only [Fig. 1(a)] and speech in noise [Fig. 1(b)]. Both data sets were extracted from an utterance of the Spanish SpeechDat-Car database[27] with uniform noise conditions so that the noise level in both data sets is the same. It can be clearly concluded that higher order statistics or polyspectra provide discriminative features for speech/nonspeech classification.[24] Even the bispectrum phase exhibits a more random behavior during nonspeech periods so that the phase entropy also provides complimentary information for VAD in noise environments.[28]

Although bispectrum methods have all the advantages of cumulants/polyspectra, their direct use has two serious limitations: (i) The computation of bispectra in the whole triangular region is huge, and (ii) the 2D template matching score in the classification is impractical. To efficiently use bispectra, integrated bispectrum methods[16,17] were proposed for different applications.[29,30]

III. INTEGRATED BISPECTRUM

Let $x(t)$ be a zero mean stationary random process. If we define $\tilde{y}(t) = x^2(t) - E\{x^2(t)\}$, the cross correlation between $\tilde{y}(t)$ and $x(t)$ is defined to be

$$r_{\tilde{y}x}(k) = E\{\tilde{y}(t)x(t + k)\} = E\{x^2(t)x(t + k)\} = C_{3x}(0,k) \quad (5)$$

so that its cross spectrum is given by

$$S_{\tilde{y}x}(\omega) = \sum_{-\infty}^{\infty} C_{3x}(0,k)\exp\{-j\omega k\} \quad (6)$$

and

$$C_{3x}(0,k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{\tilde{y}x}(\omega)\exp\{j(\omega_k)\} d\omega. \quad (7)$$
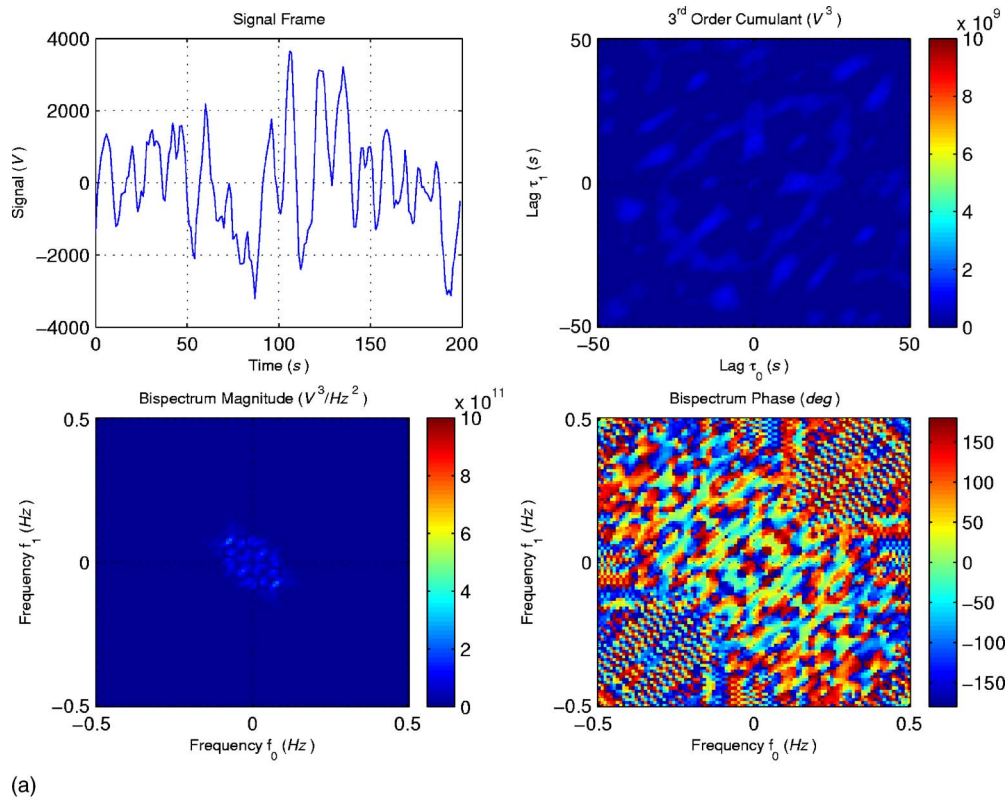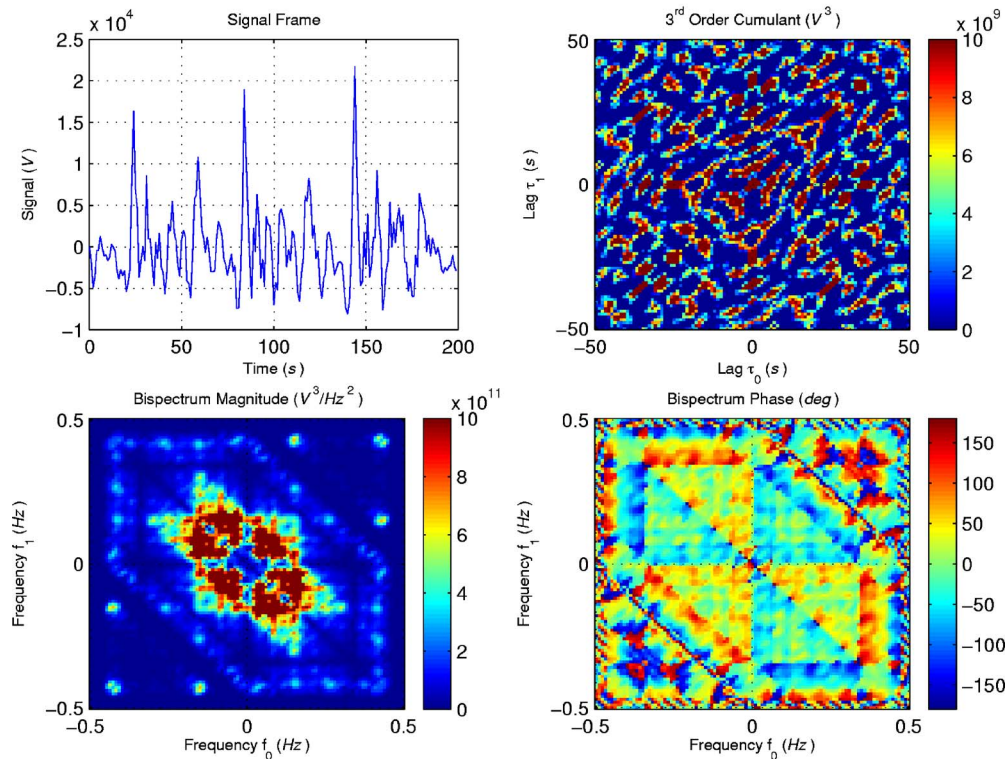
FIG. 1. Third-order statistics of: (a) a noise only signal and (b) a speech signal corrupted by car noise.

If we compare Eq. (4) with Eq. (7) we obtain

$$S_{\tilde{y}x}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} B_x(\omega, \omega_2) d\omega_2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} B_x(\omega_1, \omega) d\omega_1.$$

(8)

The integrated bispectrum is defined as a cross spectrum between the signal and its square, and therefore a function of a single frequency variable. Hence, its computation as a cross spectrum leads to significant computational savings. But more important is that the variance of the estimator is of

the same order as that of the power spectrum estimator. On the other hand, Gaussian processes have vanishing third-order moments so that the bispectrum and integrated bispectrum functions are zero as well.

Section IV shows an innovative algorithm for voice activity detection taking advantage of the statistical properties of the integrated bispectrum. The proposed method is based on a LRT that can be evaluated using a Gaussian model for the integrated bispectrum of the signal.

IV. VOICE ACTIVITY DETECTION BASED ON THE INTEGRATED BISPECTRA

This section addresses the problem of voice activity detection formulated in terms of a classical binary hypothesis testing framework:

$$
\begin{aligned}
H_0 &: \quad x(t) = n(t), \\
H_1 &: \quad x(t) = s(t) + n(t).
\end{aligned} \tag{9}
$$

In a two-hypotheses test, the optimal decision rule that minimizes the error probability is the Bayes classifier. Given an observation vector $\hat{\mathbf{y}}$ to be classified, the problem is reduced to selecting the class (H_0 or H_1) with the largest posterior probability $P(H_i|\hat{\mathbf{y}})$. From the Bayes rule a statistical LRT [8] can be defined by

$$
L(\hat{\mathbf{y}}) = \frac{p_{\mathbf{y}|H_1}(\hat{\mathbf{y}}|H_1)}{p_{\mathbf{y}|H_0}(\hat{\mathbf{y}}|H_0)}, \tag{10}
$$

where the observation vector $\hat{\mathbf{y}}$ is classified as H_1 if $L(\hat{\mathbf{y}})$ is greater than $P(H_0)/P(H_1)$, otherwise it is classified as H_0. In Ref. 22 the LRT first proposed by Sohn [8] for VAD, which was defined on the power spectrum, is generalized and extended to the case where successive observations $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_m$ of the noisy signal are available. The so-called multiple observation LRT (MO-LRT) reports significant improvements in robustness as the number of observations increases. This test involves evaluating the joint conditional distributions of the observations under H_0 and H_1,

$$
L_m(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_m) = \frac{p_{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m|H_1}(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_m|H_1)}{p_{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m|H_0}(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_m|H_0)}, \tag{11}
$$

which is easily performed if the observations are assumed to be independent.

Assuming the integrated bispectrum $\{S_{yx}(\omega):\omega\}$ as the feature vector $\hat{\mathbf{y}}$ and to be independent zero-mean Gaussian variables in the presence and absence of speech:

$$
p(S_{yx}(\omega)|H_0) = \frac{1}{\pi\lambda_0(\omega)} \exp\left[-\frac{|S_{yx}(\omega)|^2}{\lambda_0(\omega)} \right],
$$

$$
p(S_{yx}(\omega)|H_1) = \frac{1}{\pi\lambda_1(\omega)} \exp\left[-\frac{|S_{yx}(\omega)|^2}{\lambda_1(\omega)} \right], \tag{12}
$$

the evaluation of the tests defined by Eqs. (10) and (11) only requires one to estimate the integrated bispectrum $S_{yx}(\omega)$ of the noisy signal and variances λ_0 and λ_1 under absence and

presence of speech in the noisy signal. Thus, taking logarithms in Eq. (10) and substituting the model defined in Eq. (12) we obtain

$$
\begin{aligned}
\Phi(\hat{\mathbf{y}}) &= \sum_\omega \log\left(\frac{p(S_{yx}(\omega)|H_1)}{p(S_{yx}(\omega)|H_0)} \right) \\
&= \sum_\omega \left\{ \left(1 - \frac{\lambda_0(\omega)}{\lambda_1(\omega)} \right) \frac{|S_{yx}(\omega)|^2}{\lambda_0(\omega)} - \log\left(\frac{\lambda_1(\omega)}{\lambda_0(\omega)} \right) \right\}.
\end{aligned} \tag{13}
$$

Finally, if we define the *a priori* and *a posteriori* variance ratios as

$$
\xi(\omega) = \frac{\lambda_1(\omega)}{\lambda_0(\omega)} - 1, \quad \gamma(\omega) = \frac{|S_{yx}(\omega)|^2}{\lambda_0} \tag{14}
$$

Eq. (13) can be expressed in a more compact form:

$$
\begin{aligned}
\Phi(\hat{\mathbf{y}}) &= \sum_\omega \left[\left(1 - \frac{1}{1 + \xi(\omega)} \right) \gamma(\omega) - \log(1 + \xi(\omega)) \right] \\
&= \sum_\omega \left[\frac{\xi(\omega)\gamma(\omega)}{1 + \xi(\omega)} - \log(1 + \xi(\omega)) \right].
\end{aligned} \tag{15}
$$

Section IV A addresses two key issues in order to evaluate the proposed LRT:

(1) The estimation of the integrated bispectrum $S_{yx}(\omega)$ by means of a finite data set, and
(2) the computation of the variances $\lambda_0(\omega)$ and $\lambda_1(\omega)$ of the integrated bispectrum under H_0 and H_1 hypotheses.

A. Estimation of the integrated bispectrum $S_{yx}(\omega)$

Let $\hat{S}_{yx}(\omega)$ denote a consistent estimator of $S_{yx}(\omega)$ where $y(t) = x^2(t) - E\{x^2(t)\}$. Given a finite data set $\{x(1), x(2), \ldots, x(N)\}$ the integrated bispectrum is normally estimated by splitting the data set into blocks.[25] Thus, the data set is divided into K_B nonoverlapping blocks of data each of size N_B samples so that $N = K_B N_B$. Then, the cross periodogram of the ith block of data is given by

$$
\hat{S}_{yx}^{(i)}(\omega) = \frac{1}{N_B} X^{(i)}(\omega)[Y^{(i)}(\omega)]^*, \tag{16}
$$

where $X^{(i)}(\omega)$ and $Y^{(i)}(\omega)$ denote the discrete Fourier transforms of $x(t)$ and $y(t)$ for the ith block. Finally, the estimate is obtained by averaging K_B blocks,

$$
\hat{S}_{yx}(\omega) = \frac{1}{K_B} \sum_{i=1}^{K_B} \hat{S}_{yx}^{(i)}(\omega). \tag{17}
$$

This estimation is used to compute $\gamma(\omega)$ through Eq. (14).

B. Computation of the variances $\lambda_0(\omega)$ and $\lambda_1(\omega)$

The statistical properties of the bispectrum estimators have been discussed in Refs. 25 and 31. Thus, the test proposed Sec. IV A and the model assumed in Eq. (12) are justified since for large N_B, the estimates $S_{yx}^{(i)}(\omega_m)$ are complex Gaussian and independently distributed of $S_{yx}^{(i)}(\omega_n)$ for

$m \neq n(m, n = 1, 2, \ldots, N_B/2 - 1)$. Moreover, their mean and variance for large values of N_B and K_B can be approximated[16] by

$$E\{\hat{S}_{yx}(\omega)\} \approx S_{yx}(\omega),$$

$$\mathrm{var}\{\mathrm{Re}[\hat{S}_{yx}^{(i)}(\omega)]\} \approx \frac{1}{2K_B}[S_{yy}(\omega)S_{xx}(\omega) + \mathrm{Re}\{S_{yx}^2(\omega)\}],$$

$$\mathrm{var}\{\mathrm{Im}[\hat{S}_{yx}^{(i)}(\omega)]\} \approx \frac{1}{2K_B}[S_{yy}(\omega)S_{xx}(\omega) - \mathrm{Re}\{S_{yx}^2(\omega)\}],$$

$$(18)$$

where Re and Im denote the real and imaginary parts of a complex number. This means that, in order to estimate the variances λ_0 and λ_1 of the integrated bispectrum under the H_0 and H_1 hypotheses, we need to obtain an expression for $S_{xx}(\omega)$ and $S_{yy}(\omega)$ when $x(t) = n(t)$ (speech absence) and $x(t) = s(t) + n(t)$ (speech presence), respectively.

1. Speech absence

Under the hypothesis H_0, $x(t) = n(t)$ and $y(t) = x^2(t) - E\{x^2(t)\}$. Thus, $S_{xx}(\omega)$ and $S_{yy}(\omega)$ are reduced to

$$S_{xx}(\omega) = S_{nn}(\omega),$$

$$S_{yy}(\omega) = S_{n^2n^2}(\omega), \qquad (19)$$

where $S_{n^2n^2}$ can be expressed (see the Appendix) by

$$S_{n^2n^2}(\omega) = 2S_{nn}(\omega) * S_{nn}(\omega) + 2\pi\sigma_n^4 \qquad (20)$$

and λ_0 can be estimated by evaluating $S_{nn}(\omega)$ and the variance of the noise:

$$\lambda_0(\omega) = \frac{1}{K_B}[2S_{nn}(\omega) * S_{nn}(\omega) + 2\pi\sigma_n^4\delta(\omega)]S_{nn}(\omega). \quad (21)$$

2. Speech presence

Under the hypothesis H_1, $x(t) = s(t) + n(t)$ and $S_{xx}(\omega)$ and $S_{yy}(\omega)$ require a little more computation (see the Appendix):

$$S_{xx}(\omega) = S_{ss}(\omega) + S_{nn}(\omega),$$

$$S_{yy}(\omega) = S_{s^2s^2}(\omega) + S_{n^2n^2}(\omega) + 4S_{ss}(w) * S_{nn}(\omega)$$
$$- 2\pi(\sigma_s^4 + \sigma_n^4)\delta(\omega). \qquad (22)$$

By using Eq. (20) for $s(t)$ and substituting it in Eq. (22) leads to

$$S_{yy}(\omega) = 2S_{ss}(\omega) * S_{ss}(\omega) + 2S_{nn}(\omega) * S_{nn}(\omega)$$
$$+ 4S_{ss}(\omega) * S_{nn}(\omega). \qquad (23)$$

Finally, $\lambda_1(\omega)$ can be estimated in terms of $S_{ss}(\omega)$ and $S_{nn}(\omega)$ by means of

$$\lambda_1(\omega) = \frac{1}{K_B}[S_{ss}(\omega) + S_{nn}(\omega)][2S_{ss}(\omega) * S_{ss}(\omega)$$
$$+ 2S_{nn}(\omega) * S_{nn}(\omega) + 4S_{ss}(\omega) * S_{nn}(\omega)]. \qquad (24)$$
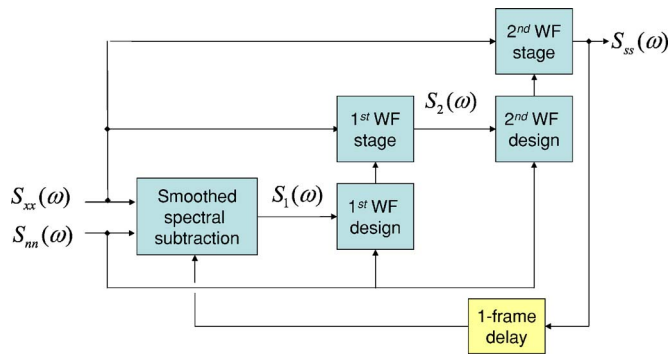
FIG. 2. Estimation of $S_{ss}(\omega)$ via smoothed spectral subtraction and Wiener filtering.

As a conclusion, a way to estimate the power spectrum of the clean signal $S_{ss}(\omega)$ is needed for the evaluation of λ_1. A method combining Wiener filtering and spectral subtraction is used in this paper to estimate $S_{ss}(\omega)$ in terms of the power spectrum $S_{xx}(\omega)$ of the noisy signal. The procedure is described as follows. During a short initialization period, the power spectrum of the residual noise $S_{nn}(\omega)$ is estimated assuming a short nonspeech period at the beginning of the utterance. Note that $S_{nn}(\omega)$ can be computed in terms of the DFT of the noisy signal $x(t) = n(t)$. After the initialization period $S_{xx}(\omega)$ is computed for each frame through Eqs. (16) and (17) and $S_{ss}(\omega)$ is then obtained by applying a denoising process. Denoising consists of a previous smoothed spectral subtraction followed by Wiener filtering. Figure 2 shows a block diagram for the estimation of the power spectrum $S_{ss}(\omega)$ of the denoised speech through the noisy signal $S_{xx}(\omega)$. It is worthwhile clarifying that $S_{nn}(\omega)$ is not only estimated during the initialization period but also updated during nonspeech frames based on the VAD decision. Thus, the denoising process consists of the following stages.

(1) Spectral subtraction:

$$S_1(\omega) = L_sS_{ss}(\omega) + (1 - L_s)\max(S_{xx}(\omega)$$
$$- \alpha S_{nn}(\omega), \beta S_{xx}(\omega)). \qquad (25)$$

(2) First WF design and filtering:

$$\mu_1(\omega) = S_1(\omega)/S_{nn}(\omega),$$

$$W_1(\omega) = \mu_1(\omega)/(1 + \mu_1(\omega)),$$

$$S_2(\omega) = W_1(\omega)S_{xx}(\omega). \qquad (26)$$

(3) Second WF design and filtering:

$$\mu_2(\omega) = S_2(\omega)/S_{nn}(\omega),$$

$$W_2(\omega) = \max(\mu_2(\omega)/(1 + \mu_2(\omega)), \beta),$$

$$S_{ss}(\omega) = W_2(\omega)S_{xx}(\omega), \qquad (27)$$

where $L_s = 0.99$, $\alpha = 1$, and $\beta = 10^{(-22/10)}$ is selected to ensure a -22 dB maximum attenuation for the filter in order to reduce the high variance musical noise that normally appears due to rapid changes across adjacent frequency bins. The main rea-
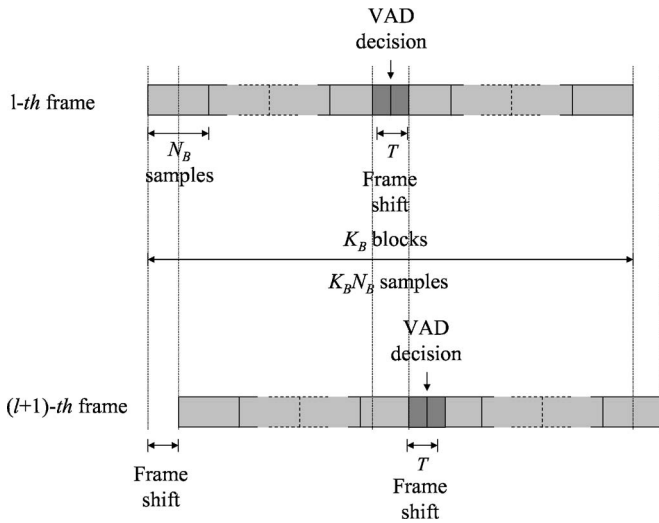
FIG. 3. Integrated bispectrum estimation by block averaging and VAD decision.

son for using a Wiener filter as noise reduction algorithm is found on its optimum performance for filtering additive noise in a noisy signal. This method is normally preferred to other conventional techniques such as spectral subtraction since the musical noise is significantly reduced as well as the variance of the residual noise. A two-stage Wiener filter configuration was used, as in the ETSI AFE standard,[13] in order to make it less sensible to the VAD decision and the noise estimation process.

V. DATA PROCESSING TECHNIQUES FOR VAD

The following shows the two different approaches for block managing that were used in this paper for the estimation of the integrated bispectrum of the input signal and its variances in the formulation of the LRT previously defined.

A. Block partitioning and averaging

The first VAD method is described as follows. The input signal $x(n)$ sampled at 8 kHz is divided into overlapping windows of size $N=K_B N_B$ samples. A typical value of the window size is about 0.2 s, which yields accurate estimations of the integrated bispectrum. The best tradeoff between block averaging (K_B) and spectral resolution (N_B) will be discussed in the next sections.

Figure 3 illustrates the way the signal is processed and the block of data the decision is made for. Note that the decision is made for a T-sample data block around the midpoint of the analysis window where T is the "frame-shift." Thus, a large data set is used to estimate the integrated bispectrum $S_{yx}(\omega)$ by averaging K_B successive blocks of data while the decision is made for a shorter data set. As in most of the standardized VADs,[3,4,13] the frame-shift is 80 samples so that the VAD frame rate is 100 Hz.

After having estimated the power spectrum $S_{ss}(\omega)$ of the clean signal through the denoising process shown earlier, the variances $\lambda_0(\omega)$ and $\lambda_1(\omega)$ of the integrated bispectrum under speech absence and speech presence are computed by evaluating the convolution operations required by Eqs. (21)

and (24), respectively. Then, the *a priori* and *a posteriori* variance ratios $\xi(\omega)$ and $\gamma(\omega)$, as defined in Eq. (14), can be estimated and the VAD decision is obtained by comparing the LRT defined in Eq. (15) to a given threshold η. If the LRT is greater than the threshold η the frame is classified as speech, otherwise it is classified as nonspeech. Finally, in order to track nonstationary noisy environments the power spectrum estimate of the noise is updated based on the current observation of the noisy signal:

$$S_{nn}(\omega) = L_n S_{nn}(\omega) + (1 - L_n) S_{xx}(\omega) \qquad (28)$$

with $L_n = 0.98$ when the VAD detects a nonspeech observation.

This method ensures a reduced variance estimation of the integrated bispectrum of the noisy signal by block averaging. However, the window shift T is typically much smaller than the block size and, therefore, the method is computationally expensive. In order to solve this problem, a computationally efficient but also effective method is developed in Sec. V B.

B. Contextual likelihood ratio test

Most VADs in use today normally consider hang-over algorithms based on empirical models to smooth the VAD decision. It has been shown recently[21,22] that incorporating long-term speech information to the decision rule reports benefits speech/pause discrimination in high noise environments, thus making unnecessary the use of hang-over mechanisms based on hand-tuned rules. The VAD previously proposed addresses this problem by formulating a smoothed decision based on a large data set. However, an optimum statistical test involving multiple and independent observations of the input signal can be also defined as in Ref. 22 over the integrated bispectrum of the noisy signal using Eq. (11).

The proposed MO-LRT formulates the decision for the central frame of a $(2m+1)$-observation buffer $\{\hat{\mathbf{y}}_{l-m}, \ldots, \hat{\mathbf{y}}_{l-1}, \hat{\mathbf{y}}_l, \hat{\mathbf{y}}_{l+1}, \ldots, \hat{\mathbf{y}}_{l+m}\}$:

$$L_{l,m}(\hat{\mathbf{y}}_{l-m}, \ldots, \cdot \hat{\mathbf{y}}_{l+m}) = \frac{p_{\mathbf{y}_{l-m}, \ldots, \mathbf{y}_{l+m}|H_1}(\hat{\mathbf{y}}_{l-m}, \ldots, \hat{\mathbf{y}}_{l+m}|H_1)}{p_{\mathbf{y}_{l-m}, \ldots, \mathbf{y}_{l+m}|H_0}(\hat{\mathbf{y}}_{l-m}, \ldots, \hat{\mathbf{y}}_{l+m}|H_0)},$$

$$\qquad (29)$$

where l denotes the frame being classified as speech (H_1) or nonspeech (H_0). Note that, assuming statistical independence between the successive observation vectors, the corresponding log-LRT:

$$\ell_{l,m} = \sum_{k=l-m}^{l+m} \ln \frac{p_{\mathbf{y}_k|H_1}(\hat{\mathbf{y}}_k|H_1)}{p_{\mathbf{y}_k|H_0}(\hat{\mathbf{y}}_k|H_0)} \qquad (30)$$

is recursive in nature, and if the Φ function is defined as

$$\Phi(k) = \ln \frac{p_{\mathbf{y}_k|H_1}(\hat{\mathbf{y}}_k|H_1)}{p_{\mathbf{y}_k|H_0}(\hat{\mathbf{y}}_k|H_0)}, \qquad (31)$$

Eq. (30) can be calculated as
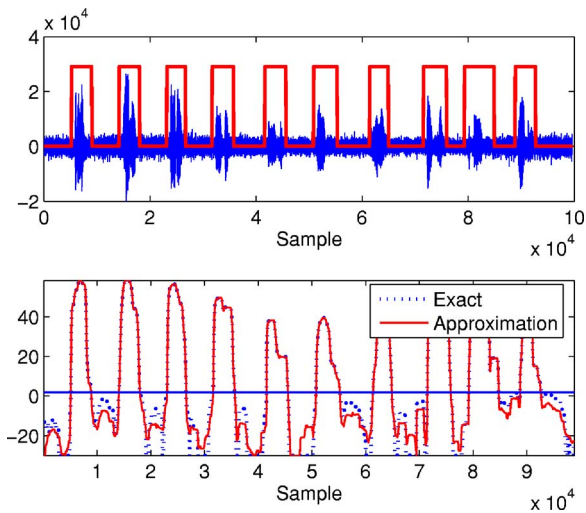
FIG. 4. Operation of the MO-LRT VAD defined on the integrated bispectrum.

$$\ell_{l+1,m} = \ell_{l,m} - \Phi(l-m) + \Phi(l+m+1). \qquad (32)$$

Now, if the integrated bispectrum of the noisy signal is considered as the feature vector, Eq. (31) is reduced to be

$$\Phi(k) = \sum_{\omega} \left[\frac{\xi_k(\omega)\gamma_k(\omega)}{1+\xi_k(\omega)} - \log(1+\xi_k(\omega)) \right], \qquad (33)$$

where $\xi_k(\omega)$ and $\gamma_k(\omega)$ denote the *a priori* and *a posteriori* variance ratios for the kth frame as defined in Eq. (14).

As a conclusion, the decision rule is formulated over a sliding window consisting of $(2m+1)$ observation vectors around the frame the decision is made for. This fact imposes an m-frame delay to the algorithm that, for several applications, including robust speech recognition, is not a serious implementation obstacle.

Figure 4 shows an example of the operation of the contextual MO-LRT VAD on an utterance of the Spanish SpeechDat-Car database.[27] Figure 4 shows the decision variables for the tests defined by Eq. (33) and, alternatively, for the test with the second log-term in Eq. (33) suppressed (*approximation*) when compared to a fixed threshold $\eta = 1.5$. For this example, $N_B = 256$ and $m = 8$. This approximation reduces the variance during nonspeech periods. It can be shown that using an eight-frame window reduces the variability of the decision variable yielding to a reduced noise variance and better speech/nonspeech discrimination. On the other hand, the inherent anticipation of the VAD decision contributes to reduce the number of speech clipping errors. The use of this test reports quantifiable benefits in speech/nonspeech detection as will be shown in Sec. VII.

C. Comparison in terms of computational cost

It is interesting to compare the two methods proposed for voice activity detection based on single and multiple-observation LRTs. Both methods exhibit the advantages of VADs employing contextual information for formulating the decision rule since the decision variable is built on a long-term data set. The first method decomposes a large analysis window into K_B blocks each of size N_B samples and the integrated bispectrum $S_{yx}(\omega)$ is estimated by averaging K_B blocks using Eq. (16) for each frame-shift T. This can be computationally expensive since the frame-shift is usually much lower than the window size ($N = K_B N_B$). The second method based on the MO-LRT is more efficient since it just requires one to compute the integrated bispectrum of the current frame (N_B samples). The test is then built on a moving average fashion by means of Eq. (32). This is clearly more efficient in terms of complexity since a single bispectrum computation is performed for each frame-shift T instead of the K_B bispectrum estimations required by the first method. On the other hand, both methods exhibit high speech/nonspeech discrimination accuracy in noisy environments for equivalent delay configurations as will be shown in the following.

VI. ANALYSIS OF THE PROPOSED METHODS

In a Bayes classifier, the overlap between the distributions of the decision variable represents the VAD error rate.[21] In order to clarify the motivations for the proposed algorithms, the distributions of the LRTs defined by Eqs. (15) and (30) were studied as a function of the design parameters K_B, N_B, and m. A hand-labeled version of the Spanish SpeechDat-Car database[27] was used in the analysis. This database contains recordings from close-talking and distant microphones at different driving conditions: (a) Stopped car, motor running, (b) town traffic, low speed, rough road, and (c) high speed, good road. The most unfavorable noise environment (i.e., high speed, good road, and distant microphone) with an average SNR of about 5 dB was selected for the experiments. Thus, the decision variables defined by Eqs. (15) and (30) were measured during speech and nonspeech periods for the whole database, and the histogram and probability distributions were built. Separate results for each of the data processing techniques discussed previously are shown in the following.

A. VAD based on block averaging

Figure 5 shows the distributions of speech and noise for different values of K_B and N_B. It is clearly shown that the distributions of speech and noise are better separated when increasing the number of blocks (K_B). When K_B increases the noise variance decreases and the speech distribution is shifted to the right being more separated from the nonspeech distribution. Thus, the distributions of speech and nonspeech are less overlapped and consequently, the error probability is reduced.

The reduction of the distribution overlap yields improvements in speech/pause discrimination. This fact can be shown by calculating the misclassification errors of speech and noise for an optimal Bayes classifier. Figure 5 also shows the areas representing the probabilities of incorrectly detecting speech and nonspeech and the optimal decision threshold. Figure 6 shows the independent decision errors for speech and nonspeech and the global error rate as a function of K_B for $N_B = 256$. The error rates were obtained by computing the areas under the class distributions of the decision variable shown in Fig. 5. The global error rate represents the
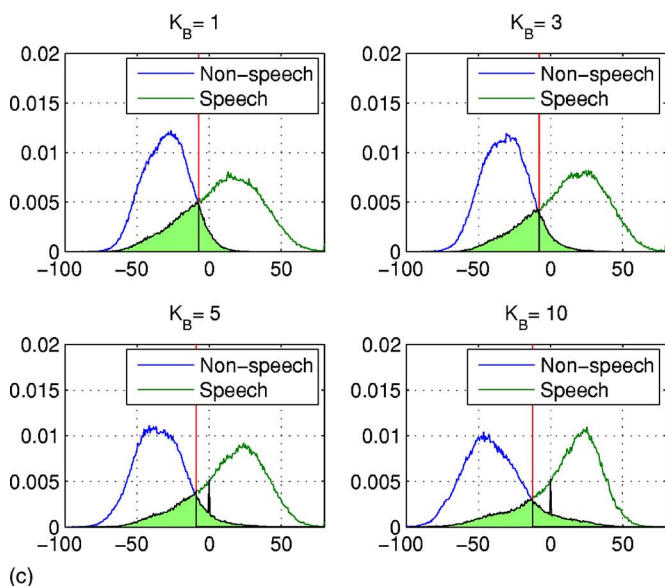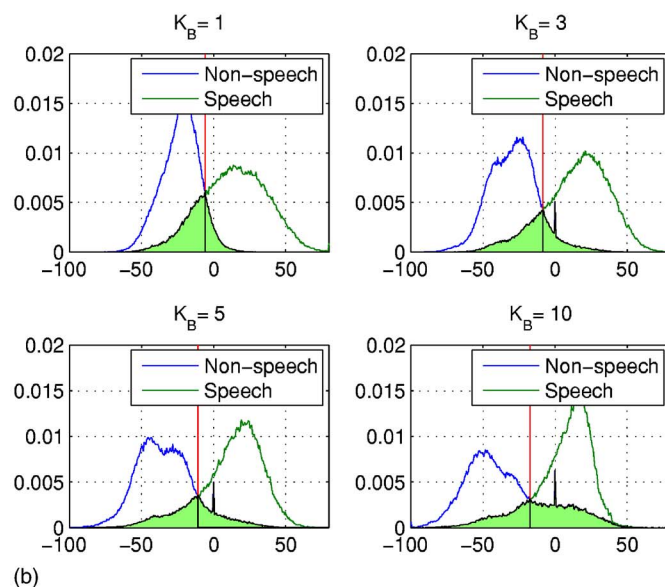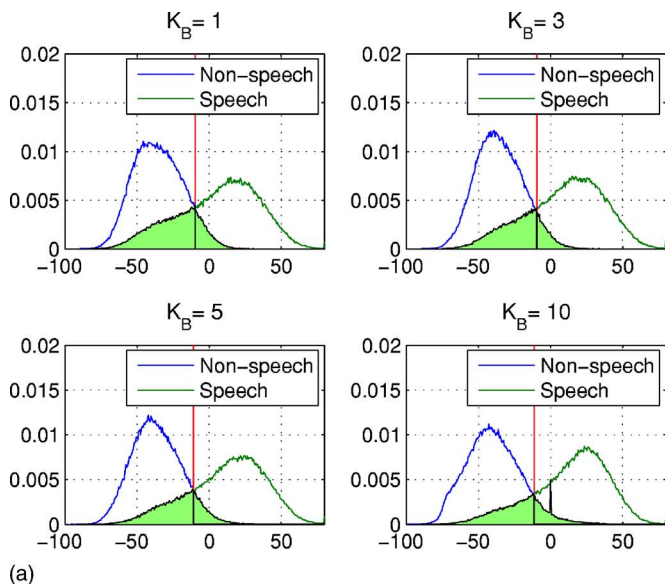
FIG. 5. Distributions of the decision variable for the VAD based on block averaging. (a) $N_B=64$. (b) $N_B=128$. (c) $N_B=256$.
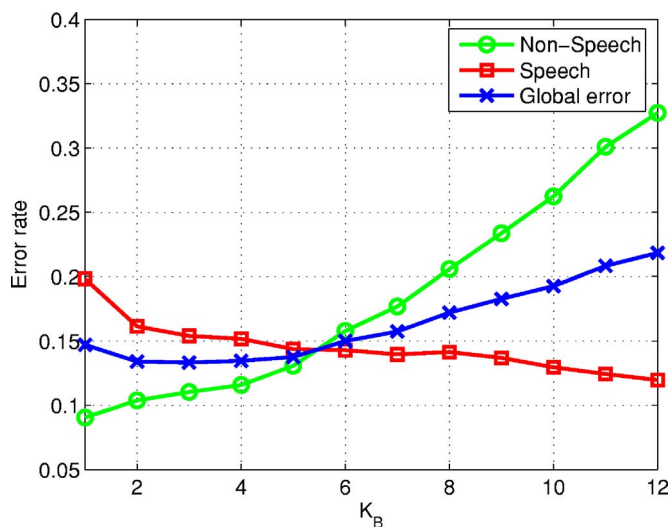
FIG. 6. Probability of error as a function of K_B for $N_B=256$.

total overlapped area while the speech and nonspeech error rate represent the areas below and above the optimum decision threshold, respectively. The speech detection error is clearly reduced when increasing the length of the window $(K_B N_B)$ while the increased robustness is only damaged by a moderate increase in the nonspeech detection error. These improvements are achieved by reducing the overlap between the distributions when K_B is increased as shown in Fig. 5. It is interesting to show that the optimal values of the parameters K_B and N_B are conditioned to a fixed window size $(K_B N_B)$. This fact is shown in Fig. 7 where the minimum value of the global error rate depends on both N_B and K_B and is obtained for a typical window size of about 640–800 samples (80–100 ms).

B. Integrated bispectrum MO-LRT VAD

Similar results were obtained for the MO-LRT VAD based on the integrated bispectrum of the noisy signal. Figure 8 shows the distributions of speech and noise for different values of K_B and $N_B=256$. Again the noise variance decreases with m and the speech distribution is shifted to the
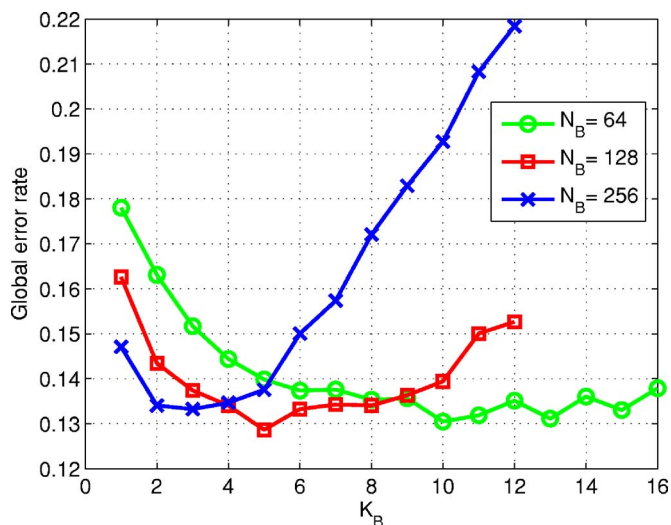
FIG. 7. Global error rates as a function of K_B for $N_B=64$, 128, and 256.

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Ramírez *et al.*: Statistical voice activity detection   2953
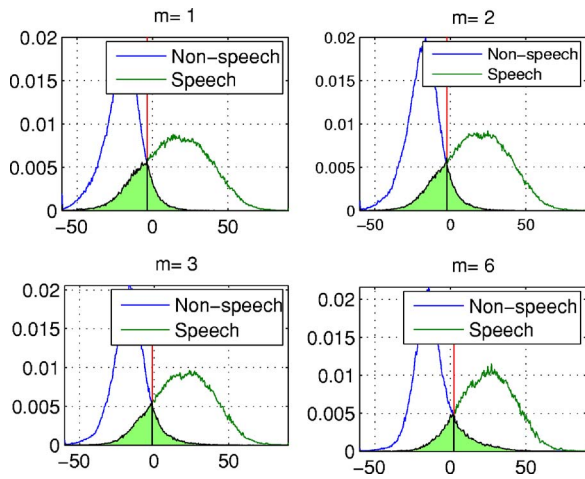
FIG. 8. Distributions of the decision variable for the VAD based on MO-LRT.

FIG. 9. Probability of error as a function of m for $N_B=256$.

right. Figure 9 shows the independent decision errors for speech and nonspeech and the global error rate as a function of m for $N_B=256$. The total error is reduced with the increasing length of the window (m) and exhibits a minimum value for a fixed order. According to Fig. 9, the optimal value of the order of the VAD is $m=6$. Thus, increasing the length of the window is beneficial in high noise environments since the VAD introduces an artificial "hang-over" period which reduces front and rear-end clipping errors. This saving period is the reason for the increase of the nonspeech detection error shown in Fig. 9.

C. Concluding remarks

The two alternative methods for data processing and definition of the decision rule yield high discrimination accuracy and minimum global classification error rate for a given length of the analysis window. It can be concluded from Figs. 7 and 9 that the best results are obtained for a typical value of the window size of about 80–100 ms. Thus, both methods benefit from using contextual information for the formulation of the decision rule. The VAD module then exhibits a delay of about half the length of the analysis window that for several applications including real-time speech transmission can be a serious implementation obstacle. However, for other applications including robust speech recognition, a delay of about 50–80 ms in the VAD module does not represent a problem and can be accepted.

VII. EXPERIMENTAL FRAMEWORK

Several experiments are commonly conducted in order to evaluate the performance of VAD algorithms. The analysis is mainly focused on the determination of the error probabilities or classification errors at different SNR levels[6] and the influence of the VAD decision on the performance of speech processing systems.[1] Subjective performance tests have also been considered for the evaluation of VADs working in combination with speech coders.[32] The follwing describes the experimental framework and the objective performance tests conducted in this paper to evaluate the proposed algorithms.
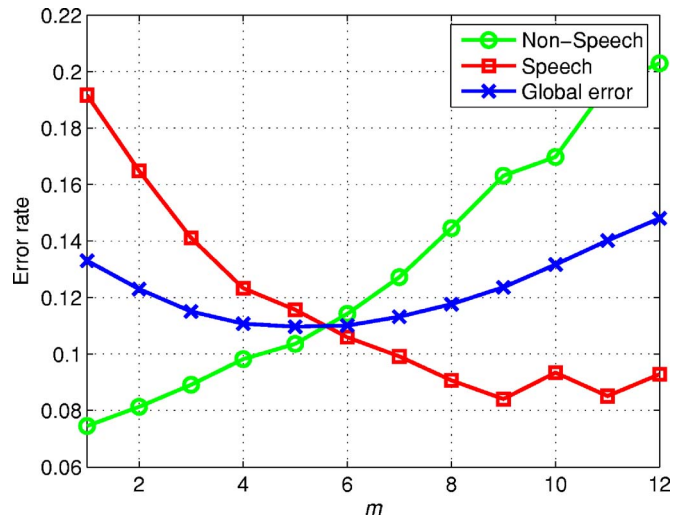
A. ROC curves

The ROC curves are frequently used to completely describe the VAD error rate. They show the tradeoff between speech and nonspeech detection accuracy as the decision threshold varies.[21] The AURORA subset of the original Spanish SpeechDat-Car database[27] was used in this analysis. This database contains 4914 recordings using close-talking and distant microphones from more than 160 speakers. The

(a)

(b)
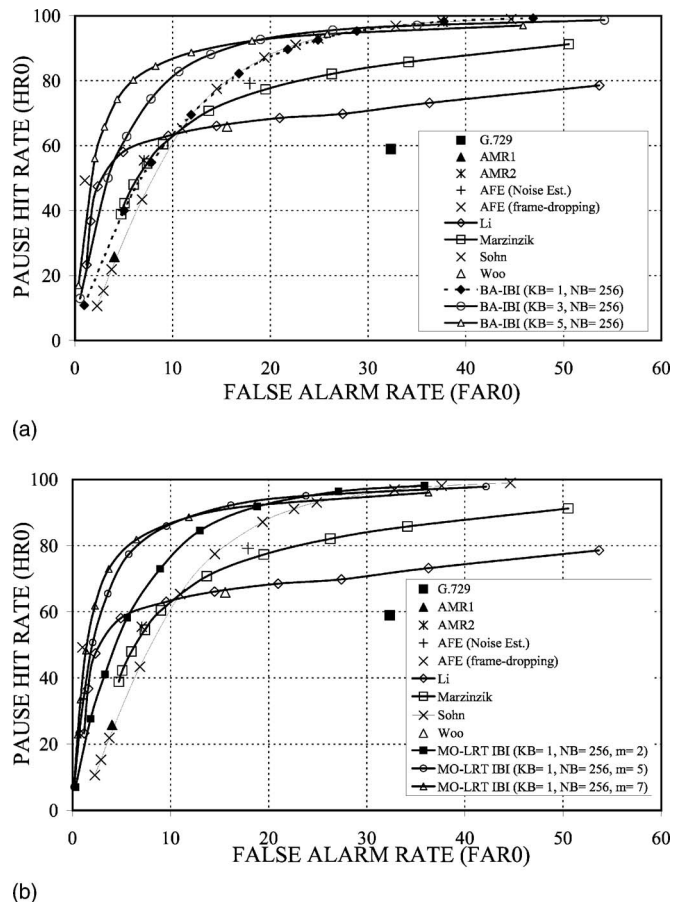
FIG. 10. ROC curves obtained in high noise conditions. (a) Block based integrated bispectrum LRT VAD. (b) Integrated bispectrum MO-LRT VAD.
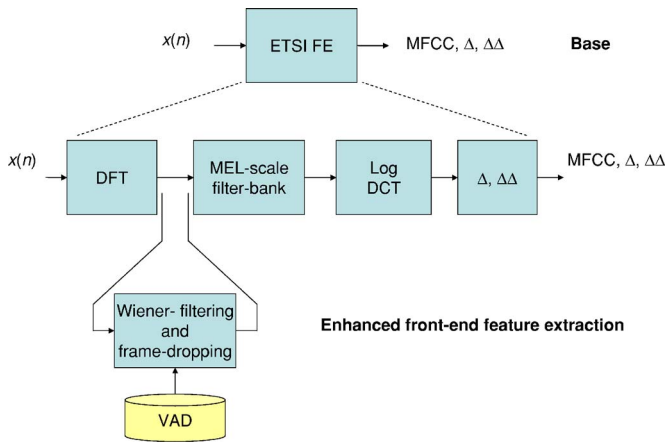
FIG. 11. Speech recognition experiments. Front-end feature extraction.

TABLE I. Average word accuracy (%) for the AURORA 2 for clean and multicondition training experiments. Results are averaged for all the noises and SNRs ranging from 20 to 0 dB.

	G.729	AMR1	AMR2	AFE	MO-LRT IBI
WF	66.19	74.97	83.37	81.57	84.15
WF+FD	70.32	74.29	82.89	83.29	85.71
	Woo	Li	Marzinzik	Sohn	Hand-labeled
WF	83.64	77.43	84.02	83.89	84.69
WF+FD	81.09	82.11	85.23	83.80	86.86

files are categorized into three noisy conditions: Quiet, low noisy, and highly noisy conditions, which represent different driving conditions with average SNR values between 25 and 5 dB. The nonspeech hit rate (HR0) and the false alarm rate (FAR0=100-HR1) were determined for each noise condition being the actual speech frames and actual speech pauses determined by hand-labeling the database on the close-talking microphone.

Figure 10 shows the ROC curves of the proposed VADs and other frequently referred algorithms[6,8,14,15] for recordings from the distant microphone in high noisy conditions. The working points of the G.729, AMR and AFE VADs are also included. Figure 10(a) shows how increasing the number of blocks (K_B) in the block averaging integrated bispectrum (BA-IBI) LRT VAD leads to a shift-up and to the left of the ROC curve in the ROC space. This result is consistent with the analysis conducted in Figs. 5 and 7 that predicts a minimum error rate for K_B close to five blocks. Similar results are obtained for the efficient MO-LRT IBI VAD that exhibits a shift of the ROC curve when the number of observations (m) increases as shown in Fig. 10(b). Again, the results are consistent with our preliminary experiments and the results shown in Figs. 8 and 9 that expect a minimum error rate for m close to eight frames. Both methods show clear improvements in detection accuracy over standardized VADs and over a representative set of recently published VAD algorithms.[6,8,14,15]

Thus, among all the VADs examined, our VAD yields the lowest false alarm rate for a fixed nonspeech hit rate, and also the highest nonspeech hit rate for a given false alarm rate. The benefits are especially important over G.729, which is used along with a speech codec for discontinuous transmission, and over the algorithm of Li et al. (Ref. 15), that is based on an optimum linear filter for edge detection. The proposed VAD also improves Marzinzik VAD[6] that tracks the power spectral envelopes, and the Sohn VAD[8] that formulates the decision rule by means of a statistical likelihood ratio test defined on the power spectrum of the noisy signal.

It is worthwhile mentioning that the above-described experiments yield a first measure of the performance of the VAD. Other measures of VAD performance that have been reported are the clipping errors.[32] These measures provide valuable information about the performance of the VAD and can be used for optimizing its operation. Our analysis does not consider or analyze the position of the frames within the word and assesses the hit rates and false alarm rates for a first performance evaluation of the proposed VAD. On the other hand, the speech recognition experiments conducted later on the AURORA databases will be a direct measure of the quality of the VAD and the application it was designed for. Clipping errors are indirectly evaluated by the speech recognition system since there is a high probability of a deletion error to occur when part of the word is lost after frame-dropping.

B. Speech recognition experiments

Although the ROC curves are effective for VAD evaluation, the influence of the VAD in a speech recognition system was also studied. Many authors claim that VADs are well compared by evaluating speech recognition performance[14] since nonefficient speech/nonspeech classification is an important source of the degradation of recognition performance in noisy environments.[2] There are two clear motivations for that: (i) Noise parameters such as its spectrum are updated during nonspeech periods being the speech enhancement system strongly influenced by the quality of the noise estimation, and (ii) FD, a frequently used technique in speech recognition to reduce the number of insertion errors caused by the noise, is based on the VAD decision and speech misclassification errors lead to loss of speech, thus causing irrecoverable deletion errors. This section evaluates the VAD according to the objective it was developed for, that is, by assessing the influence of the VAD in a speech recognition system.

TABLE II. Average word accuracy (%) for the Spanish SDC databases.

	Base	Woo	Li	Marz.	Sohn	G.729	AMR1	AMR2	AFE	MO-LRT IBI
WM	92.94	95.35	91.82	94.29	96.07	88.62	94.65	95.67	95.28	**96.39**
MM	83.31	89.30	77.45	89.81	91.64	72.84	80.59	90.91	90.23	**91.75**
HM	51.55	83.64	78.52	79.43	84.03	65.50	62.41	85.77	77.53	**86.65**
Avg.	75.93	89.43	82.60	87.84	90.58	75.65	74.33	90.78	87.68	**91.60**

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Ramírez et al.: Statistical voice activity detection   2955

Figure 11 shows a block diagram of the speech recognition experiments conducted to evaluate the proposed VAD. The reference framework (base) considered for these experiments is the ETSI AURORA project for distributed speech recognition.[33] The recognizer is based on the HTK (Hidden Markov Model Toolkit) software package.[34] The task consists of recognizing connected digits which are modeled as whole word HMMs (Hidden Markov Models) with 16 states per word, simple left-to-right models, and three Gaussian mixtures per state (diagonal covariance matrix). Speech pause models consist of three states with a mixture of six Gaussians per state. The 39-parameter feature vector consists of 12 cepstral coefficients (without the zero-order coefficient), the logarithmic frame energy plus the corresponding delta and acceleration coefficients. Two training modes are defined for the experiments conducted on the AURORA-2 database: (i) Training on clean data only (Clean Training), and (ii) training on clean and noisy data (Multi-Condition Training). For the AURORA-3 SpeechDat-Car databases, the so called well-matched (WM), medium-mismatch (MM), and high-mismatch (HM) conditions are used. These databases contain recordings from the close-talking and distant microphones. In WM condition, both close-talking and hands-free microphones are used for training and testing. In MM condition, both training and testing are performed using the hands-free microphone recordings. In HM condition, training is done using close-talking microphone recordings from all the driving conditions while testing is done using the hands-free microphone at low and high noise driving conditions. Finally, recognition performance is assessed in terms of the word accuracy (WAcc) that considers deletion, substitution, and insertion errors.

An enhanced feature extraction scheme incorporating a noise reduction algorithm and nonspeech FD was built on the base system.[33] The noise reduction algorithm has been implemented as a single WF stage as described in the AFE standard[13] but without mel-scale warping. No other mismatch reduction techniques already present in the AFE standard have been considered since they are not affected by the VAD decision and can mask the impact of the VAD precision on the overall system performance.

Table I shows the recognition performance achieved by the different VADs that were compared. These results are averaged over the three test sets (A, B, and C) of the AURORA-2 recognition experiments[35] and SNRs between 20 and 0 dB. Note that, for the recognition experiments based on the AFE VADs, the same configuration of the standard,[13] which considers different VADs for WF and FD, was used. The proposed integrated bispectrum MO-LRT VAD outperforms the standard G.729, AMR1, AMR2, and AFE VADs in both clean and multicondition training/testing experiments. When compared to recently reported VAD algorithms, the proposed one yields better results being the one that is closer to the "ideal" hand-labeled speech recognition performance.

Table II shows the recognition performance for the Spanish SpeechDat-Car database when WF and FD are performed on the base system.[33] Again, the VAD outperforms all the algorithms used for reference yielding relevant improvements in speech recognition. Note that these particular databases used in the AURORA 3 experiments have longer nonspeech periods than the AURORA 2 database, and then the effectiveness of the VAD results are more important for the speech recognition system. This fact can be clearly shown when comparing the performance of the proposed VAD to Marzinzik VAD.[6] The word accuracy of both VADs is quite similar for the AURORA 2 task. However, the proposed VAD yields a significant performance improvement over Marzinzik VAD[6] for the AURORA 3 database.

VIII. CONCLUSIONS

This paper showed two different schemes for improving speech detection robustness and the performance of speech recognition systems working in noisy environments. Both methods are based on statistical likelihood ratio tests defined on the integrated bispectrum of the speech signal which is defined as a cross spectrum between the signal and its square, and inherits the ability of higher order statistics to detect signals in noise with many other additional advantages: (i) Its computation as a cross spectrum leads to significant computational savings, and (ii) the variance of the estimator is of the same order of power spectrum estimators. The proposed methods incorporate contextual information to the decision rule, a strategy that has reported significant improvements in speech detection accuracy and robust speech recognition applications. They differ in the way the signal is processed in order to obtain precise estimations of the integrated bispectrum and its variance. The optimal window size was determined by analyzing the overlap between the distributions of the decision variable and the error rate of an optimum Bayes classifier. The experimental analysis conducted on the well-known AURORA databases has reported significant improvements over standardized techniques such as ITU G.729, AMR1, AMR2, and ESTI AFE VADs, as well as over recently published VADs. The analysis assessed: (i) The speech/nonspeech detection accuracy by means of the ROC curves, with the proposed VAD yielding improved hit rates and reduced false alarms when compared to all the reference algorithms, and (ii) the recognition rate when the VAD is considered as part of a complete speech recognition system, showing a sustained advantage in speech recognition performance.

APPENDIX: VARIANCES OF THE INTEGRATED BISPECTRUM FUNCTION

This appendix demonstrates Eqs. (20) and (22) that were used in Sec. IV B for the computation of the variances $\lambda_0(\omega)$

and $\lambda_1(\omega)$ of the integrated bispectrum function under speech absence and speech presence, respectively.

Let us assume the clean signal $s(t)$ and the noise $n(t)$ to be stationary, zero mean $(E[s(t)]=E[n(t)]=0)$, statistically independent processes so that

$$r_x(k) = r_s(k) + r_n(k) \Rightarrow S_{xx}(\omega) = S_{ss}(\omega) + S_{nn}(\omega), \quad \text{(A1)}$$

where $r_x(k)$ denotes the autocorrelation function of the signal corrupted by additive noise.

In order to derive the variances of the integrated bispectrum under the hypotheses H_0 and H_1, it is first needed to evaluate the autocorrelation function of the sequence $y(t) = x^2(t) - E[x^2(t)]$, which is defined as

$$r_{yy}(k) = E[y(t)y(t+k)]$$
$$= E[(x^2(t) - E[x^2(t)])(x^2(t+k) - E[x^2(t+k)])]. \quad \text{(A2)}$$

If the variance of the signal is defined as $\sigma_x^2 = E[x^2(t)] = E[x^2(t+k)]$ then

$$r_{yy}(k) = E[(x^2(t)x^2(t+k)] - \sigma_x^4. \quad \text{(A3)}$$

Moreover, if the clean signal and the noise are assumed to be noncorrelated $(E[s(t)n(t)]=E[s(t)]E[n(t)]=0)$ the second term on the right-hand side of Eq. (A3) can be expressed as

$$\sigma_x^4 = E^2[(s(t) + n(t))^2] = (\sigma_s^2 + \sigma_n^2 + \overbrace{2E[s(t)n(t)]}^{0})^2 = \sigma_s^4$$
$$+ \sigma_n^4 + 2\sigma_s^2\sigma_n^2. \quad \text{(A4)}$$

By defining $\bar{y}(t) \equiv x^2(t)$, Eq. (A3) can be expressed as

$$r_{\bar{y}\bar{y}}(k) \equiv E[(s(t) + n(t))^2(s(t+k) + n(t+k))^2]$$
$$= E[s^2(t)s^2(t+k)] + E[n^2(t)n^2(t+k)]$$
$$+ E[s^2(t)n^2(t+k)] + 2E[s^2(t)n(t+k)s(t+k)]$$
$$+ E[n^2(t)s^2(t+k)] + 2E[n^2(t)n(t+k)s(t+k)]$$
$$+ 2E[n(t)s(t)s^2(t+k)] + 2E[n(t)s(t)n^2(t+k)]$$
$$+ 4E[n(t)s(t)n(t+k)s(t+k)], \quad \text{(A5)}$$

where $r_{s^2s^2} = E[s^2(t)s^2(t+k)]$ and $r_{n^2n^2} = E[n^2(t)n^2(t+k)]$.

Using the relation between fourth-order moments and cumulants given in Ref. 36 for a set or random variables:

$$C_{x_1,\dots,x_n} = \sum_{p_1,\dots,p_m} (-1)^{m-1}(m-1)!$$
$$\times E\left[\prod_{i \in p_1} X_i\right] \cdots E\left[\prod_{i \in p_m} X_i\right], \quad \text{(A6)}$$

where $\{p_1, \dots, p_m\}$ are all the partitions with $m = 1, \dots, n$ of the set of integers $\{1, \dots, r\}$. In particular, for $p=4$:

$$C_{x_1,x_2,x_3,x_4} = (-1)^0(0!)E[x_1x_2x_3x_4] + (-1)^1(1!)$$
$$\times[E[x_1x_2]E[x_3x_4] + E[x_1x_4]E[x_2x_3]$$
$$+ E[x_1x_3]E[x_2x_4] + E[x_1]E[x_2x_3x_4]$$
$$+ E[x_2]E[x_1x_3x_4] + E[x_3]E[x_1x_2x_4]$$

$$+ E[x_4]E[x_1x_2x_3]] + (-1)^2(2!)$$
$$\times[\times E[x_1]E[x_2]E[x_3x_4] + E[x_1]E[x_3]E[x_2x_4]$$
$$+ E[x_1]E[x_4]E[x_2x_3] + E[x_2]E[x_3]E[x_1x_4]$$
$$+ E[x_2]E[x_4]E[x_1x_3] + E[x_3]E[x_4]E[x_1x_2]]$$
$$+ (-1)^3(3!)E[x_1]E[x_2]E[x_3]E[x_4]. \quad \text{(A7)}$$

Note that Eq. (A7) can be significantly reduced if the random variables are assumed to be zero mean:

$$C_{x_1,x_2,x_3,x_4} = E[x_1x_2x_3x_4] - [E[x_1x_2]E[x_3x_4]$$
$$+ E[x_1x_4]E[x_2x_3] + E[x_1x_3]E[x_2x_4]]. \quad \text{(A8)}$$

Under the statistical independency assumption, the cross cumulants are null and the cross terms in Eq. (A5) are reduced to

$$E[s^2(t)n^2(t+k)] = \overbrace{C_{s,s,n,n}}^{0} + 2E[s(t)n(t+k)]E[s(t)n(t+k)]$$
$$+ E[s^2(t)]E[n^2(t+k)],$$

$$2E[s^2(t)n(t+k)s(t+k)]$$
$$= 2(\overbrace{C_{s,s,n,s}}^{0} + 2E[s(t)n(t+k)]E[s(t)s(t+k)]$$
$$+ E[s^2(t)]E[s(t+k)n(t+k)]),$$

$$E[n^2(t)s^2(t+k)] = \overbrace{C_{n,n,s,s}}^{0} + 2E[n(t)s(t+k)]E[n(t)s(t+k)]$$
$$+ E[n^2(t)]E[s^2(t+k)],$$

$$2E[n^2(t)n(t+k)s(t+k)]$$
$$= 2(\overbrace{C_{n,n,s}}^{0} + 2E[n(t)n(t+k)]E[n(t)s(t+k)]) + E[n^2(t)]E[s(t+k)n(t+k)]),$$

$$2E[n(t)s(t)s^2(t+k)]$$
$$= 2(\overbrace{C_{n,s,s,s}}^{0} + 2E[n(t)s(t+k)]E[s(t)s(t+k)]$$
$$+ E[n(t)s(t)]E[s^2(t+k)]),$$

$$2E[n(t)s(t)n^2(t+k)]$$
$$= 2(\overbrace{C_{n,s,n,n}}^{0} + 2E[n(t)n(t+k)]E[s(t)n(+k)]$$
$$+ E[n(t)s(t)]E[n^2(t+k)]),$$

$$4E[n(t)s(t)n(t+k)s(t+k)]$$
$$= 4(\overbrace{C_{n,s,n,s}}^{0} + E[n(t)n(t+k)]E[s(t)s(t+k)]$$
$$+ E[n(t)s(t+k)]E[s(t)n(t+k)]$$
$$+ E[s(t)s(t+k)]E[n(t)n(t+k)]).$$

$$\text{(A9)}$$

Using these expressions, Eq. (A5) is reduced to

$$r_{\bar{y}\bar{y}}(k) = r_{s^2s^2}(k) + r_{n^2n^2}(k) + 2\sigma_s^2\sigma_n^2 + r_s(k)r_n(k) \quad \text{(A10)}$$

and Eq. (A3) can be expressed as

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Ramírez et al.: Statistical voice activity detection   2957

$$r_{yy}(k) = r_{s^2s^2}(k) + r_{n^2n^2}(k) + 4r_s(k)r_n(k) - (\sigma_s^4 + \sigma_n^4).$$
$$(A11)$$

Finally, the power spectrum of the squared centered signal $y(t)$ is obtained after computing the discrete Fourier transform (DFT) on Eq. (A11):

$$S_{yy}(\omega) = S_{s^2s^2}(\omega) + S_{n^2n^2}(\omega) + 4S_{ss}(\omega) * S_{nn}(\omega)$$
$$- 2\pi(\sigma_s^4 + \sigma_n^4)\delta(\omega), \qquad (A12)$$

where the asterisk "$*$" denotes convolution in the frequency domain.

Finally, under speech absence $x(t) = n(t)$ and assuming $n(t)$ to be a Gaussian process:

$$E[n^2(t)n^2(t+k)] = \overbrace{C_{n,n,n,n}}^{0} + 2\overbrace{E[n(t)n(t+k)]E[n(t)n(t+k)]}^{r_n(k)r_n(k)}$$
$$+ \overbrace{E[n^2(t)]E[n^2(t+k)]}^{\sigma_n^4}$$
$$(A13)$$

or equivalently:

$$S_{n^2n^2}(\omega) = 2S_{nn}(\omega) * S_{nn}(\omega) + 2\pi\sigma_n^4\delta(\omega). \qquad (A14)$$

Note that, using the above-derived equations, the variances of the integrated bispectrum function under H_1 and H_0 hypotheses can be computed and the statistical tests used in the proposed VAD methods are completely described and justified.

[1] R. L. Bouquin-Jeannes and G. Faucon, "Study of a voice activity detector and its influence on a noise reduction system," Speech Commun. **16**, 245–254 (1995).

[2] L. Karray and A. Martin, "Towards improving speech detection robustness for speech recognition in adverse environments," Speech Commun. 261–276 (2003).

[3] ETSI, "Voice activity detector (VAD) for Adaptive Multi-Rate (AMR) speech traffic channels," ETSI EN 301 708 Recommendation, 1999 (European Telecommunications Standards Inst., France).

[4] ITU, "A silence compression scheme for G.729 optimized for terminals conforming to recommendation V.70," ITU-T Recommendation G.729-Annex B, 1996 (International Telecomm. Union, Geneva).

[5] A. Sangwan, M. C. Chiranth, H. S. Jamadagni, R. Sah, R. V. Prasad, and V. Gaurav, "VAD techniques for real-time speech transmission on the Internet," in IEEE International Conference on High-Speed Networks and Multimedia Communications, 2002, pp. 46–50.

[6] M. Marzinzik and B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," IEEE Trans. Speech Audio Process. **10**, 341–351 (2002).

[7] D. K. Freeman, G. Cosier, C. B. Southcott, and I. Boyd, "The voice activity detector for the pan-european digital cellular mobile telephone service," in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 1989, pp. 369–372.

[8] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," IEEE Signal Process. Lett. **16**, 1–3 (1999).

[9] I. Potamitis and E. Fishler, "Speech activity detection and enhancement of a moving speaker based on the wideband generalized likelihood ratio and microphone arrays," J. Acoust. Soc. Am. **116**, 2406–2415 (2004).

[10] J. Górriz, J. Ramírez, J. C. Segura, and C. Puntonet, "An effective cluster-based model for robust speech detection and speech recognition in noisy environments," J. Acoust. Soc. Am. **120**, 470–481 (2006).

[11] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 1979, pp. 208–211.

[12] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120 (1979).

[13] ETSI, "Speech processing, transmission and quality aspects (STQ); Distributed speech recognition; advanced front-end feature extraction algorithm; Compression algorithms," ETSI ES 202 050 Recommendation, 2002 (European Telecommunications Standards Inst., France).

[14] K. Woo, T. Yang, K. Park, and C. Lee, "Robust voice activity detection algorithm for estimating noise spectrum," Electron. Lett. **36**, 180–181 (2000).

[15] Q. Li, J. Zheng, A. Tsai, and Q. Zhou, "Robust endpoint detection and energy normalization for real-time speech and speaker recognition," IEEE Trans. Speech Audio Process. **10**, 146–157 (2002).

[16] J. K. Tugnait, "Detection of non-Gaussian signals using integrated polyspectrum," IEEE Trans. Signal Process. **42**, 3137–3149 (1994).

[17] J. K. Tugnait, "Corrections to detection of non-Gaussian signals using integrated polyspectrum," IEEE Trans. Signal Process. **43**, 2792–2793 (1995).

[18] J. Ramírez, J. C. Segura, M. C. Benítez, A. de la Torre, and A. Rubio, "A new adaptive long-term spectral estimation voice activity detector," in Proceedings of EUROSPEECH 2003, Geneva, Switzerland, pp. 3041–3044.

[19] A. Sangwan, W. Zhu, and M. Ahmad, "Improved voice activity detection via contextual information and noise suppression," in IEEE International Symposium on Circuits and Systems (ISCAS), 2005, pp. 868–871.

[20] J. Ramírez, J. C. Segura, C. Benítez, A. de la Torre, and A. Rubio, "An effective subband osf-based vad with noise reduction for robust speech recognition," IEEE Trans. Speech Audio Process. **13**, 1119–1129 (2005).

[21] J. Ramírez, J. C. Segura, M. C. Benítez, A. de la Torre, and A. Rubio, "Efficient voice activity detection algorithms using long-term speech information," Speech Commun. **42**, 271–287 (2004).

[22] J. Ramírez, J. C. Segura, C. Benítez, L. García, and A. Rubio, "Statistical voice activity detection using a multiple observation likelihood ratio test," IEEE Signal Process. Lett. **12**, 689–692 (2005).

[23] J. Górriz, J. Ramírez, J. Segura, and C. Puntonet, "Improved MO-LRT VAD based on bispectra Gaussian model," Electron. Lett. **41**, 877–879 (2005).

[24] J. Ramírez, J. M. Górriz, J. C. Segura, C. G. Puntonet, and A. Rubio, "Speech/non-speech discrimination based on contextual information integrated bispectrum LRT," IEEE Signal Process. Lett. **13** (2006).

[25] D. R. Brillinger and M. Rosenblatt, *Spectral Analysis of Time Series* (Wiley, New York, 1968).

[26] C. Nikias and M. Raghuveer, "Bispectrum estimation: A digital signal processing framework," Proc. IEEE **75**, 869–891 (1987).

[27] A. Moreno, L. Borge, D. Christoph, R. Gael, C. Khalid, E. Stephan, and A. Jeffrey, "SpeechDat-Car: A large speech database for automotive environments," in Proceedings of the II LREC Conference 2000.

[28] J. M. Górriz, J. Ramírez, C. G. Puntonet, and J. Segura, "An efficient bispectrum phase entropy-based algorithm for VAD," in Interspeech 2006, pp. 2322–2325.

[29] X. Zhang, Y. Shi, and Z. Bao, "A new feature vector using selected bispectra for signal classification with application in radar target recognition," IEEE Trans. Signal Process. **49**, 1875–1885 (2001).

[30] X. Liao and Z. Bao, "Circularly integrated bispectra: Novel shift invariant features for high-resolution radar target recognition," Electron. Lett. **34**, 1879–1880 (1998).

[31] D. Brillinger, *Time Series Data Analysis and Theory* (Holt, Rinehart and Winston, New York, 1975).

[32] A. Benyassine, E. Shlomot, H. Su, D. Massaloux, C. Lamblin, and J. Petit, "ITU-T Recommendation G.729 Annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications," IEEE Commun. Mag. **35**, 64–73 (1997).

[33] ETSI, "Speech processing, transmission and quality aspects (stq); distributed speech recognition; front-end feature extraction algorithm; compression algorithms," ETSI ES 201 108 Recommendation, 2000 (European Telecommunications Standards Inst., France).

[34] S. Young, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, *The HTK Book* (Cambridge University Press, New York, 1997).

[35] H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noise conditions," in ISCA ITRW ASR2000 Automatic Speech Recognition: Challenges for the Next Millennium, Paris, France, 2000 (Intl. Speech Communication Assn.).

[36] C. Nikias and A. Petropulu, *Higher Order Spectra Analysis: a Non-linear Signal Processing Framework* (Prentice Hall, Englewood Cliffs, NJ, 1993).

Loudness control in pianists as exemplified in keystroke force measurements on different touches

Hiroshi Kinoshita[a) and Shinichi Furuya
Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Tomoko Aoki
Faculty of Environmental and Symbiotic Sciences, Prefectural University of Kumamoto,
3-1-100 Kumamoto City, Kumamoto 862-8502, Japan

Eckart Altenmüller
Institute of Music Physiology and Musicians' Medicine, University of Music and Drama Hanover,
Hohenzollernstrasse 47 D-30161 Hannover, Germany

The relationship between the key depression force on an upright piano and the level of loudness of a generated tone was examined when pianists hit a force-sensor built-in key with "struck" or "pressed" type of touch. The vertical displacement of the key, and the radiated piano sounds were also recorded. It was found that for both types of touch, simple exponential functions could adequately describe the relation of the force amplitude with the level of the piano tone as well as that of the impulse of the force with the piano tone. The impulse of the force generated before the maximum key depression moment commonly amounted to above 80% of the total impulse produced at the tone below *mezzo-forte*. It, however, decreased to around 60% at *fortissimo*, indicating a decrease in the efficiency of the force application for sound production. The two types of touch differed in their force profiles. The struck touch was characterized by a steeper initial force increase with greater fluctuations in the subsequent period than the pressed touch. The struck touch also demonstrated lower maximum force and less impulse at *fortissimo*. The inter-pianist variation in the force and impulse, and the "finger-noise" are also herein examined. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2717493]

I. INTRODUCTION

Playing the piano undoubtedly represents one of the most complex sensori-motor skills that humans may acquire, requiring years of training. Striking the key to generate the required level of loudness is a fundamental and indispensable component of a pianist's skill. The keystroke commonly starts with a downswing of the arm toward the key, and ends with the application of regulated depression force by the fingertip on the front side of the key. The key itself moves in a "see-saw"-type of motion via a balance-rail bearing as the player presses the key. The applied downward force thus acts on a complex hammer-driving structure resting on the other side of the key via an assembly called a "whippen jack." Shortly before the hammer contacts the string, the transmission of force from the key to the hammer is interrupted and the hammer is left swinging freely against the string, allowing an effective transfer of kinetic energy from the hammer to the string for sound production (Fletcher and Rossing, 1998).

An understanding of the force or impulse acting on the key by the finger is important because it is the sole source of kinetic energy to move the hammer. Hence, it can be a major variable for pianists to modulate the volume or pressure of the generated sound. When considering the nature of sound in addition, the manner in which pianists apply force to the key must also be included as a variable. Two different fundamental techniques of key touch may be of interest in this respect. One technique is the so-called "struck touch," where the key is struck by a moving finger, and another is the "pressed touch," when the key is pressed by a finger resting on its surface (Goebl *et al.*, 2005). The force generated between the finger and the key ("finger force") can also be regarded as a source of stress on the pianist's hand as it may reach a relatively large magnitude immediately after the moment of key-front rail contact (Harding *et al.*, 1989). Repetitive application of such stress is considered to be one of the major causes of musculoskeletal problems in keyboard players (Amadio and Russotti, 1990; Caldron *et al.*, 1986; Fry, 1991).

Some efforts have been made to study the force applied by the finger on the key (finger force) in the past. Using a thin force sensor foil inserted into the space between the key and front rail in the key bed, Parlitz *et al.* (1998) measured the depressing force of the finger against the front rail while performing so-called "tied-finger exercises," i.e., some fingers were depressing piano keys and holding them down for the entire exercise, while the remaining fingers executed the key stroke. Groups of expert and amateur players performed three tied-finger exercises with increasing degrees of difficulty while the loudness level was kept constant at around a

forte (*f*) level. They found that the amateurs used greater tied-finger forces as well as keystroke forces, and a longer force application period by the striking fingers compared with the experts on playing the same notes. They concluded that years of piano training allow for pianists with an independent coordination of playing and nonplaying fingers after touch and a sense for the piano's response.

Using a force transducer mounted on the key surface, Harding *et al.* (1989) directly measured finger force at a moderate loudness level during *staccato* and *legato* strikes by subjects with different skill levels. The finger force after the moment when the key reaches its maximum displacement ("bottom out") cannot contribute to sound production. Therefore, they computed the value of impulse (the area under the force-time curve) generated prior to the bottom-out moment. Since they did not measure key displacement, they assumed that the bottom-out moment corresponded to the onset moment of the highest peak force. The findings indicated that this impulse value did not differ between the *staccato* and *legato* strikes when producing the same level of loudness. The impulse values had a nonlinear relationship with key velocity. The force profiles shown in their study also suggested that the impulse generated from the key touch to the bottom-out moment was smaller than the impulse generated after the bottom-out moment.

Using a similar force measurement system, Askenfelt and Jansson (1992) measured the finger force at various loudness levels. The number of pianists as well as the force sensor type used was not mentioned in their article, and the description of the force characteristics was limited. Nevertheless, they provided important information about the magnitude of maximum finger force in relation to sound loudness. They stated that at *piano* (*p*), *mezzo forte* (*mf*), and *fortissimo* (*ff*) levels with *staccato* touch, the peak force reached around 8, 15, and 50 N, respectively. In *legato* touch, the finger forces were considerably lower during acceleration, typically one third or even less at the softest levels.

It is apparent from these reports that little work has been done to investigate the nature of force imparted by the finger on the piano key. Especially, detailed finger-force information based on a larger number of pianists is lacking. It is also necessary to have a key in which a force sensor is built in and the sensor surface is flush with the key surface so that any psychological factor influencing the control of keystroke movement is minimized. In the present study, therefore, using a force transducer built-in key, we made a more complete analysis of the relationship between the finger force and level of sound pressure (SPL) using ten expert pianists who performed keystrokes with struck and pressed touches.

II. METHODS

A. Subjects

Eight female and two male active classical pianists (age=21.6±1.7 yrs., height=163.6±8.8 cm, weight =52.1±10.1 kg) who showed no serious playing-related physical problems in the past served as the subjects in the present study. Each pianist had at least 15 years of training
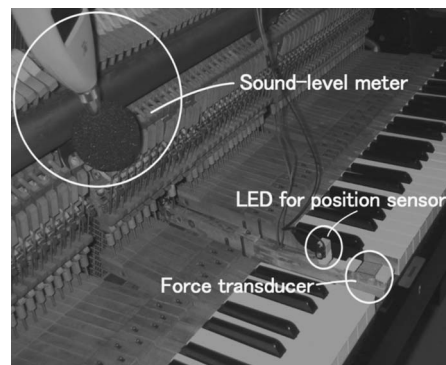
FIG. 1. Experimental setup.

with experience of winning a prize(s) at high-level domestic (Japan) and/or international piano competitions. Informed consent was obtained from each subject prior to the experiment, and the study was approved by the Ethics Committee for Human Research at Osaka University.

B. Finger force, key displacement and sound data acquisition methods

The finger force was measured using a strain-gauge-type miniature uniaxial force transducer fixed to the distal end of a C4 key of a Yamaha upright U1-model piano (see the "force transducer" in Fig. 1). The transducer was designed and built for the purpose of this study based on our previous work (Kinoshita *et al.*, 1995). Therefore, the measurement range of 0–100 N within a 0.5% error in linearity was chosen. The resolution of the transducer was 0.02 N. The natural frequency of the unloaded force transducer was DC-1 kHz. The force signal was amplified using a Kyowa strain gauge amplifier and stored on a SONY personal computer via a 12 bit analog-to-digital (A/D) converter sampling at a frequency of 900 Hz. The vertical movement of the key was also recorded using a Hamamatsu photosonic light-spot two-dimensional position measurement system interfaced with a personal computer. The light-spot LED was placed 1.5 cm above the key surface and 8 cm from the front edge of the key (see the "LED for position sensor" in Fig. 1). The key displacement data recorded were smoothed at a cut-off frequency of 24 Hz using a second-order Butterworth digital filter. Vertical velocity and acceleration were then calculated using a numerical differentiation method. Three adjacent keys (B4, C4, and D4) were removed from the piano to assure the free space for the transducer wires.

During the experiment, radiated piano sounds were synchronically sampled using a RION sound-level meter placed 20 cm above the key (see the "sound-level meter" in Fig. 1), which was then amplified using an audio amplifier to feed into the personal computer via an A/D converter sampling frequency at 900 Hz.

The experimental room was an ordinary temperature-controllable room. The background noise level was around 50 dB. This relatively louder background noise level was due to numerous machines in the experimental room.

2960   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Kinoshita *et al.*: Sound pressure and keystroke force

C. Key-striking task

The key-striking task was a series of slow- and self-paced (freq.=approx. 0.3 Hz) repetitive keystrokes with very short duration. The length of each tone generated was instructed to be a 16th note without holding each tone, and thus in a *presto* tempo. For each subject, data were collected from a series of 60 strokes at varied SPLs for each of the two prototypical types of touch: depressing the keys with the finger initially resting on the key surface (pressed) and hitting the key from a certain distance above (struck) (Askenfelt and Jansson, 1991; Goebl *et al.*, 2004; 2005). To obtain SPLs distributed approximately evenly over the whole range between the maximum and minimum, the following instruction was given to each subject prior to the series for each touch: "During the initial 15 strokes, please increase the level of sound generated gradually and voluntarily from the minimum to the maximum level, and during the following 15 strokes, please decrease the generated sound gradually to the minimum level. Please repeat the same process during the following 30 strokes. Please count the number of strokes silently for each of the 15 strokes." At the end of the experiment with each subject, the collected and stored data were checked to have a total of 120 keystrokes (60 struck and 60 pressed).

As an initial hand position common to the struck and pressed touches, each subject was instructed to place his/her right middle finger by lightly (>0.03 N) touching its tip on the C4 key while the tips of the thumb and little finger were also lightly touching the A4 and E4 keys, respectively. The index and ring fingers were kept in the air because the B4 and D4 keys were not present. The left arm and hand were relaxed and kept on the left thigh during the experiment.

For the struck touch, from the initial hand position, the right arm was lifted to a self-determined distance above the key surface at self-determined natural speed. From this lifted position, the arm was dropped and the C4 key was struck by the right middle finger to generate piano sound. After this, the whole arm was immediately lifted again as a follow through to a self-determined height. The hand was then returned to the starting position at the self-determined speed to prepare for the next keystroke.

For the pressed touch, there was no arm lift prior to a keystroke. The C4 key was therefore pressed by the right middle finger to generate sound from the initial hand position without any preparatory arm lift. After the keystroke, however, the whole arm was immediately lifted as a follow through to a self-determined height similar to the case of the struck touch. After this, the hand was lowered to the starting position to prepare for the next keystroke.

Prior to the experiment, each subject practiced these experimental tasks until they felt comfortable performing them.

D. Parameters evaluated

For the computation of sound pressure, the peak amplitude of the absolute value of the sampled sound data as a voltage was used (Palmer and Brown, 1991). The voltage value (V) was then converted to a SPL representation in dB using the following equation: $\text{SPL} = L_0 + 10\log_{10}$
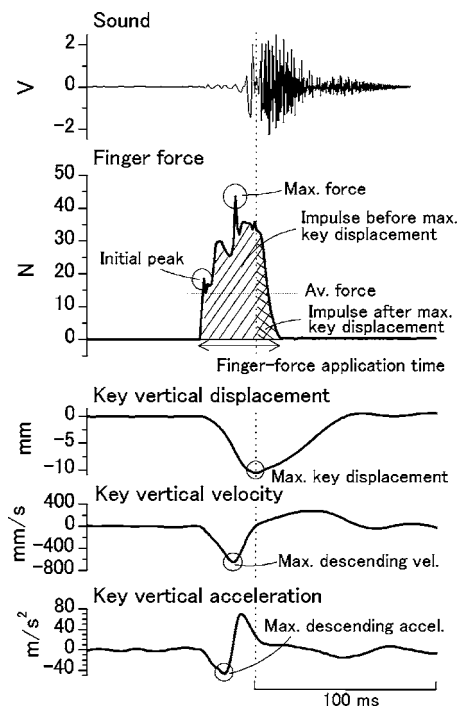
FIG. 2. Representative sound, finger-force, key-displacement, key-velocity, and key-acceleration curves during striking the key with a struck touch. Some of the parameters evaluated are also shown.

$(V^2/(V_0/\sqrt{2})^2)$, where L_0 and V_0 indicate the dB value and the corresponding voltage value of fundamental sound, respectively.

Nine variables describing the finger force and kinematics of the key were computed from each trial data set for each subject, and used for the evaluation in the present study (Fig. 2). These were: (1) the finger-force application time as defined by the duration between the onset (<0.06 N) and termination (>0.06 N) of finger-force application for the keystroke, (2) initial peak force, (3) maximum force, (4) average force during the finger-force application time, (5) total impulse as defined by an integration of the force during the finger-force application time, (6) impulse over the period between the onset of finger force for the keystroke and maximum displacement of the key, (7) impulse over the period from the maximum key displacement to the termination of the finger force, (8) maximum displacement of the key, and (9) maximum descending acceleration of the key. The maximum key descending velocity was also computed, but it was not used for evaluation in the present study.

III. RESULTS

A. Finger force profiles

Figure 3 shows typical examples of time history curves for the sound signal, finger force, and key displacement from one subject with the struck (the upper panel) and pressed (the lower panel) types of key touch at three different SPLs (95, 103, and 111 dBs). These SPLs corresponded to typical *pp*, *mf*, and *ff* levels with the present piano according to the subjective judgment of the present subjects.

The force profiles for the struck touch (see also Fig. 2) were characterized by a rapid rise in force immediately after

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Kinoshita *et al.*: Sound pressure and keystroke force   2961
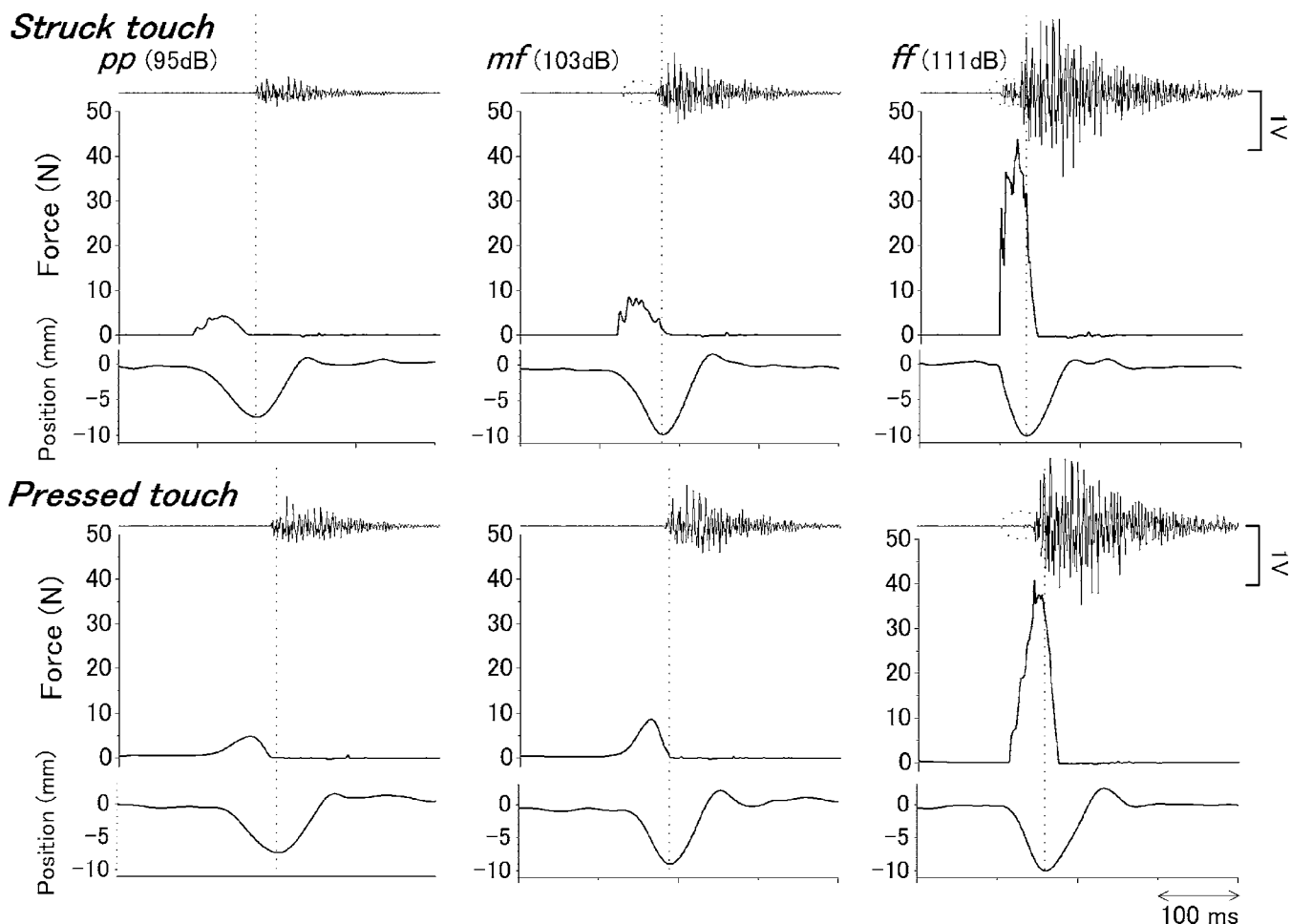
FIG. 3. Representative profiles of finger force, key displacement, and sound from selected trials (*pp*, *mf*, and *ff*) of one subject with struck (upper panel) and pressed (lower panel) types of key touch. The dotted line indicates the moment of maximum key displacement. The touch noise is indicated by a dashed circle in the sound signal.

the finger contacted the key, which produced an initial peak at 3–5 ms after contact. The force then rapidly dropped to show a distinct trough after another 3–5 ms from the peak. The force was then developed again, and formed multiple peaks until the end of the finger-force application time. The sound signals indicated that at all SPLs, the piano sound was generated after the appearance of the maximum finger force. The key displacement curves demonstrated that the key started to move soon after the finger force was developed, and it reached the lowest position toward the end of the finger-force application time.

The force during a pressed touch developed more slowly than that during a struck touch (lower panel in Fig. 3). The initial peak of the force was less clear with the pressed touch, though a small step-like initial force increase was noted in some cases, especially when generating louder sound (see the force curve at *ff*). The subsequent force with the pressed touch was also less variable than that with the struck touch. Similar to the struck touch, the piano sound was generated after the appearance of the maximum finger force, and also the moment of the maximum key displacement occurred near the end of the finger-force application time.

B. Maximum force, total impulse, and finger-force application time in relation to SPL

Figure 4 shows typical examples of the maximum force, average force, total impulse, and finger-force application time as a function of SPL generated by one subject during the struck (upper panel) and pressed (lower panel) types of key touch. For both touch types, the forces and impulses increased curvilinearly with an increased SPL, while the finger-force application time decreased linearly with SPL. An attempt was made to fit simple mathematical functions to describe the curvilinear relationship of the force and impulse parameters of each subject. It was found that an exponential function ($F = A \exp(B * \mathrm{SPL}) + C$, where F = finger force, A, B, and C are constants) was sufficient to describe the relationship rather precisely compared with six other functions tested. With this curve fitting, the proportions of variance (r^2) computed for the maximum force for each subject ranged from 0.81 to 0.98 (mean ± SD = 0.91 ± 0.05), and those for the total impulse were from 0.83 to 0.96 (mean ± SD = 0.88 ± 0.04). The corresponding values for the pressed touch were from 0.87 to 0.96 (mean ± SD = 0.93 ± 0.03) for the maximum force, and from 0.75 to 0.96 (mean ± SD = 0.86 ± 0.09) for the total impulse. The finger-force applica-

Struck touch
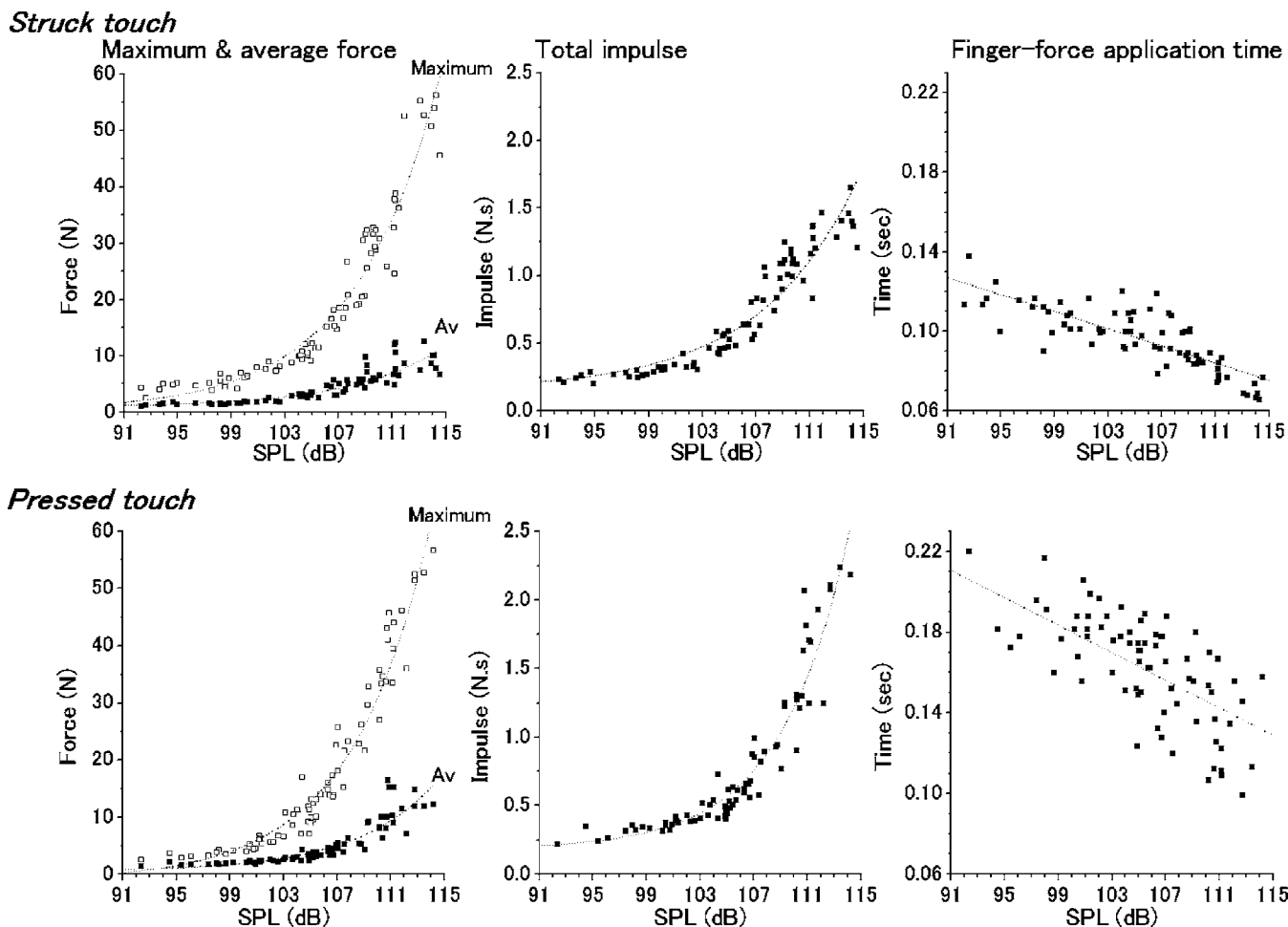
Pressed touch

FIG. 4. Changes in the maximum force, average force, total impulse, and finger-force application time with SPL of one subject with the struck touch (upper panel), and the pressed touch (lower panel). The line in the figure indicates a fit curve for the data. The data are from the same subject in Fig. 3.

tion time decreased linearly with SPL in all subjects for the struck touch (mean $r^2 \pm SD = 0.79 \pm 0.12$) and the pressed touch (mean $\pm SD = 0.68 \pm 0.09$).

A simple exponential function also fitted relatively well with the maximum force and total impulse data pooled across all subjects (Fig. 5). The r^2 values computed for the maximum force were 0.78 for the struck touch and 0.81 for the pressed touch. The r^2 values computed for the total impulse data were 0.66 for both types of key touch.

C. Impulse before and after maximum key displacement

Since the piano sound is generated by finger force to move the key downward, the impulse of the force produced during the period of downward movement should reflect the physical effort of the pianists for the production of an intended SPL. Any force applied to the key after the attainment of maximum key displacement may then be wasted in terms of sound production. Impulses generated both before and after maximum key displacement were computed for each keystroke for all subjects. In Figs. 6(A) and 6(B), the before- and after-values of the impulse are plotted in relation to SPL for both types of key touch. With an increase of SPL, the before- and after-values were increased in a curvilinear fashion.

The efficiency of finger force action was assessed by computing a proportion of the before value of the impulse to the total impulse produced. The computed value decreased from nearly 100% at lower SPLs to around 65% at higher SPLs with means and SDs of $82.5 \pm 9.0\%$ for the struck touch and $80.4 \pm 12.2\%$ for the pressed touch (Figs. 6(C) and 6(D)).

D. Comparison between the struck and pressed touch modes

Finger force was compared between the two types of key touch at different SPLs. For this purpose, the SPL data between 93 and 113 dBs were arbitrarily sorted into five categories roughly corresponding to musical dynamics indication with each being 3.99 dB wide. Within each of these dB categories, the mean values for the maximum force, impulse before the maximum key displacement, and maximum acceleration were computed for each subject, and subsequently for all subjects (Fig. 7). At each SPL category, a "planned comparison" using one-way analysis of covariance (ANOVA) with repeated measures was performed for each of these force and acceleration variables as a dependent variable and the touch mode as an independent variable (see, the explanation about the "planned comparison" in Keppel, 1991). There was a significantly higher maximum force ($F_{1,9}$
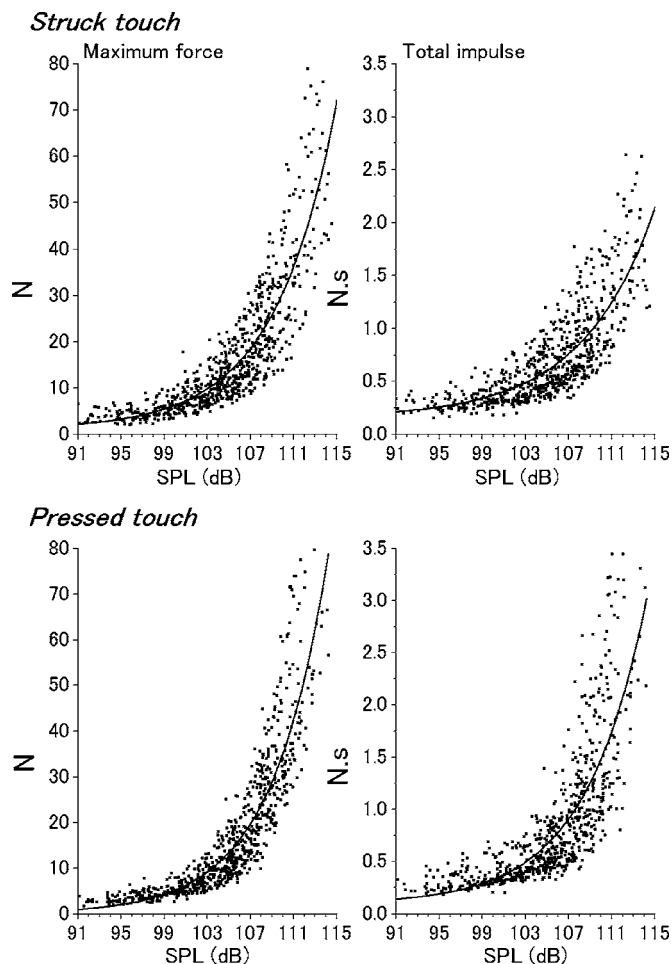
FIG. 5. Changes in the maximum force and total impulse with SPL of all subjects for the struck (upper panel) and pressed (lower panel) types of key touch. The line fit was made using a simple exponential function.

FIG. 6. Impulse before and after maximum key displacement in relation to SPL for all subjects with struck (A) and pressed (B) types of key touch, and percent values of impulse before the maximum key displacement relative to the total impulse for the struck (C) and pressed (D) types of key touch.

$=6.79$, $p < 0.05$) and impulse ($F_{1,9}=7.07$, $p < 0.05$), and a smaller descending acceleration ($F_{1,9}=18.3$, $p < 0.001$) for the pressed touch than for the struck touch at the highest SPL range (109–113 dB) (Fig. 7). At the lower SPL range (97–100.99 dB), on the other hand, the struck touch had a higher maximum force ($F_{1,9}=10.78$, $P < 0.01$).

E. Inter-subject variation in finger force

In Fig. 8(A), maximum forces are plotted separately for all ten pianists playing with the pressed touch only. Fit curves of the force- or impulse-SPL relationship exhibited a clear variation in the absolute value of applied force among the pianists especially when producing louder sound. To assess inter-subject variation at each of the dynamic categories as defined above, the coefficient of variation values (CV in %) for the maximum force and impulse were computed by dividing the SD by the corresponding mean. The CVs ranged between 27 and 31% across all ranges of SPL without showing any increasing or decreasing trend with SPL.

Why are the finger forces of individual subjects different when simply striking the same key and producing the same level of sound? One explanation for this may be because of the difference in pianists' body mass; heavy pianists with a larger upper body mass strike the key with greater force. We

therefore examined the relation between the maximum finger force and body weight of the subjects for each of the above-defined five SPL ranges. In Fig. 8(B), the relations for the 93–96.99, 101–104.99, and 109–113 dB ranges are shown as examples. Note that the correlation values computed were quite small, and they were all statistically insignificant (see Fig. 8(B)). The r values for the 97–100.99 and 105–108.99 dB ranges were −0.067 and −0.209, respectively. Plots for these data are not shown in Fig. 8(B) because they could overcrowd the figure.

Although the entire mass of the body is not a related factor of the maximum finger force, another concept of mass, which is the "effective mass," is yet to be examined since it reflects a more directly related portion of the body concerning the keystroke. The effective mass can be defined here as "a portion of the pianist body (mostly the upper extremity) that is accelerated to generate the external force," and thus it is estimated by the measured force divided by its measured acceleration. A relationship between the maximum finger force and the ratio between the maximum force and maximum key acceleration as an estimate of the effective mass was thus examined (Fig. 8(C)). Significantly high correlations were found at all of the SPL ranges (see the r values of the 93–96.99, 101–104.99, and 109–113 dB ranges in Fig. 8(C)). The r values for the 97–100.99 and 105–108.99 dB ranges were 0.696 ($p < 0.05$) and 0.920 ($p < 0.005$), respec-
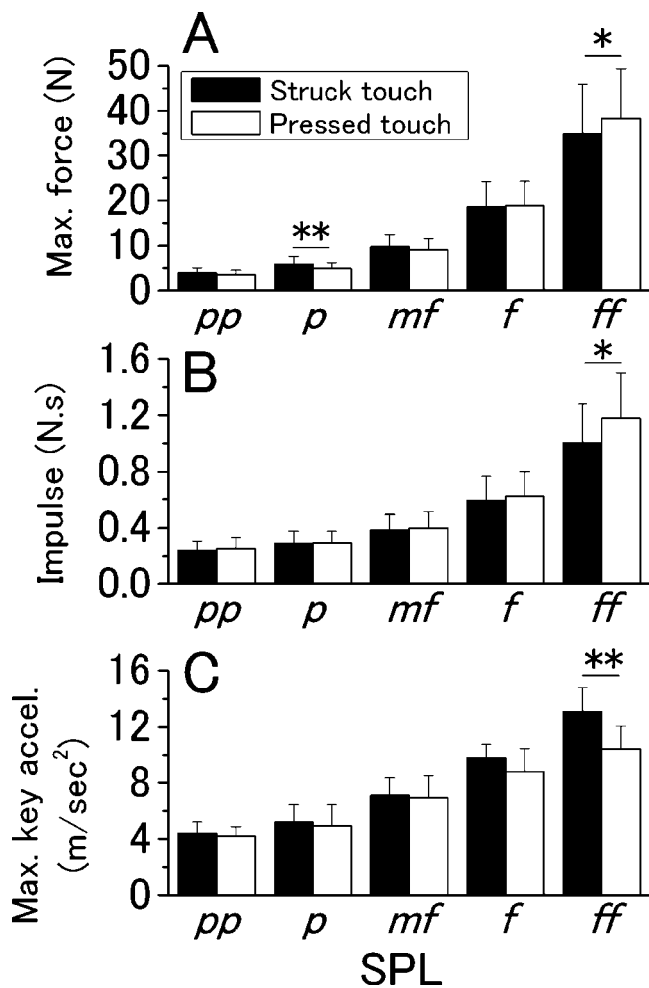
FIG. 7. A comparison between keystrokes with struck and pressed types of key touch for the maximum force (A), impulse before maximum key displacement (B), and acceleration of the key during the descending phase (C) for different dynamic categories (pp: 93–96.99 dB, p: 97–100.99 dB, mf: 101–104.99 dB, f: 105–108.99 dB, ff: 109–113 dB). A planned comparison using one-way repeated measures ANOVA was performed at each of the SPL ranges. $^*p < 0.05$. $^{**}p < 0.01$. Error bars represent +1SDs.

tively. Therefore, from 49 to 87% of the variance for the maximum finger force may be explained by the effective mass. We also examined the relationship between the maximum force and maximum key acceleration, which was insignificant at each of the SPL ranges (Fig. 8(D), $r = -0.100$ at 97–100.99 dB and -0.128 at 105–108.99 dB).

F. Finger-touching noise

The sound signals indicated that at all SPLs the piano sound was generated after the appearance of the maximum finger force. On the other hand, there was commonly a small but detectable sound signal starting immediately after the finger-key contact moment for the struck touch (see dotted circles in the sound signals of mf and ff of the struck touch in Fig. 3; also Fig. 2), and after a short period of force application for the pressed touch (see "ff" of the pressed touch in Fig. 3). This first soft sound has been termed "finger noise" or "touch precursor" (Goebl *et al.*, 2005), and is caused mostly by the finger or nail when it pounds or rubs the key surface. Using criteria of >59.0 dB for the SPL and >5 mm for the key position, we automatically detected and differentiated this noise from the background noise as well as the piano sound. For the struck touch, the onset of this signal coincided temporally with the occurrence of the initial finger-force peak.

A clear touch-related noise was detected in 560 keystrokes (93%) among the 600 keystrokes performed by ten subjects for the struck touch. The noise occurred 40.1 ± 18.2 ms (mean ± SD of the 560 observations) before the emergence of the piano sound. The noise SPL increased markedly from 62 to 65 dB at the initial peak force of 2–3 N to 90–100 dB at the peak force of 30–45 N (Fig. 9(A)). The piano SPLs corresponding to these finger noises were 90–95 dB and 110–115 dB, respectively. The difference between the SPLs of the finger noise and piano sound were therefore about 30 dB at pp-p levels and about 15 dB at f-ff.
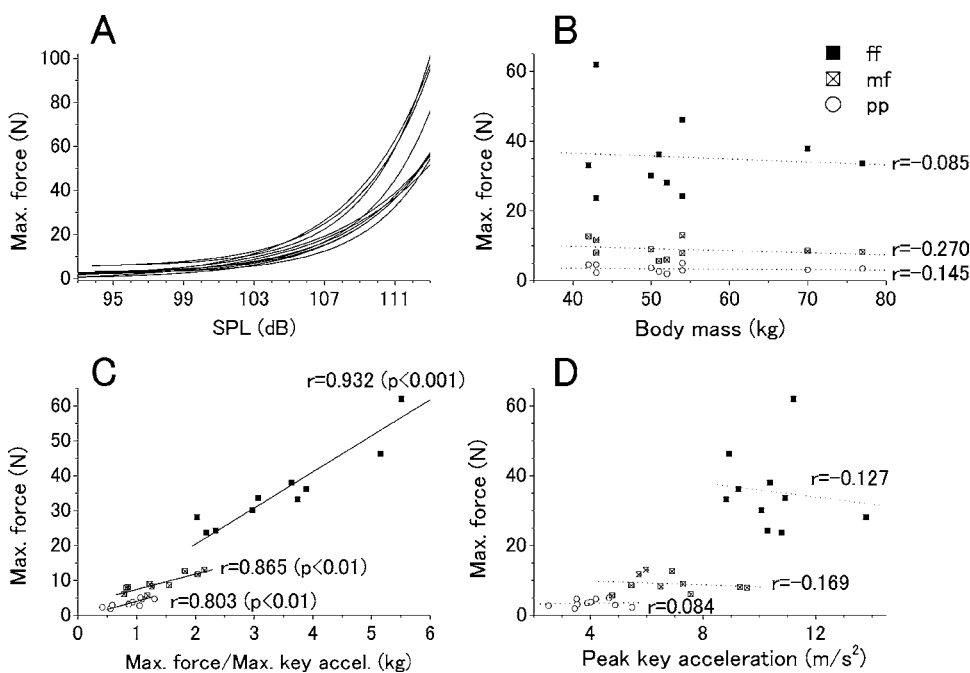
FIG. 8. Inter-subject variation in the maximum force-SPL relationship curves for all subjects with the pressed touch (A), and relationships between the maximum force and body mass (B), between the maximum force and an estimate of the effective mass (maximum force/maximum acceleration) (C), and between the maximum force and peak acceleration of the key (D). The maximum force and maximum acceleration in B, C, and D are the average values for each subject computed at pp, mf, and ff of the SPL ranges defined in Fig. 7. The data were computed only for the pressed touch. Lines indicate the regression lines computed.
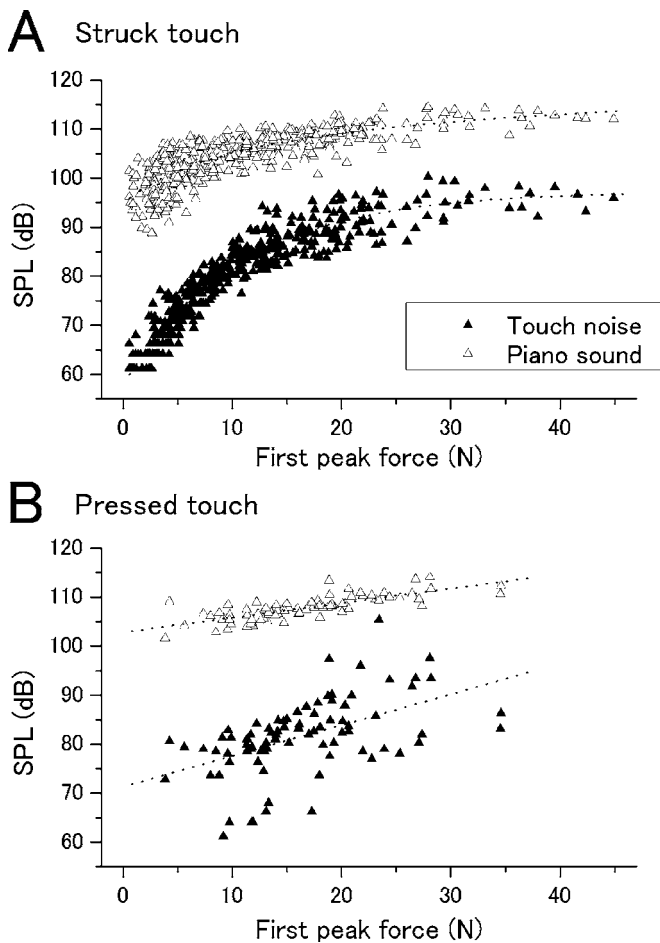
FIG. 9. Finger noise and related piano sound as a function of the first peak of finger force with the struck (A) and pressed (B) types of key touch. The data plotted are from all subjects. The dotted lines indicate best fit curves with a logistic function for the struck touch and a linear function for the pressed touch.

For the pressed touch, the finger noise-like signal was detected in 163 keystrokes (44% of 600 keystrokes for all subjects) from the automatic analysis of the data. Since the finger is not pounding the key during the pressed touch, we consider that it may not be safe to categorize these sound signals in the same way as the noise detected with the struck touch; it may be totally key-action-related mechanical noise. Of 163 keystrokes, 110 occasions had a detectable initial first peak of finger force, and therefore we tentatively plotted those in relation to the initial peak force (Fig. 9(B)). The noises ranged from 61 to 106 dB, which occurred 23.9±7.3 ms (mean ± SD) before the emergence of the piano sound. The SPL difference between the piano sound and finger noise was about 25 dB for the pressed touch.

IV. DISCUSSION

A. Finger force at differing SPLs

The maximum finger forces observed in the expert pianists commonly ranged between 3 N at *pp* and 60 N at *ff*. Four of them also exceeded 60 N at *ff* or above. This force range was wider than those (8 N at *p* to 50 N at *ff*) reported by Askenfelt and Jansson (1992) due possibly to our slightly wider SPL range examined. However, it may also be attrib-

uted to differences in the ease of hitting the force sensor which is hidden under the key surface as in the present study, in contrast to a positioning on the top of the key in the latter study. A higher required level of accuracy in hitting the sensor (especially at *ff* in the latter setting) therefore could have caused a reduction in the attacking velocity.

Our results indicate that a simple exponential function can describe adequately the changes in the parameters of finger force in relation to SPL regardless of the key touch type. The increase of the maximum finger force with an increase in SPL, and therefore the force/SPL ratio, was quite small at low SPLs. However, it became very large at high SPLs. For example, at *pp* (from 93 to 96.99 dB), the ratio was around 0.4 N/dB, whereas at *ff* (from 109 to 113 dB), it became 6.5 N/dB for the struck touch and 8 N/dB for the pressed touch, showing a 16- to 20-fold increase. The adjustment of low sound intensities clearly demanded an extremely high level of force control. Conversely, at the highest SPLs, this demands the modulation of a large force output against the key. How are pianists coping with these contrasting demands? Our previous studies of the upper extremity movements of expert pianists who performed keystrokes at various SPLs demonstrated that the production of small SPLs was accomplished principally by exclusive movements of the fingers while keeping the proximal limbs relatively stable (Furuya *et al.*, 2006). The production of higher SPLs was, on the other hand, reached by increasing both the range and speed of joint movements in the proximal limbs. In addition, there was an increased activation level of the flexor-extensor muscle pairs of the whole upper limb at the moment of key contact. These findings suggested that SPL control was made principally by modulating the mass of the body portions involved in key depression, joint stiffness, and the attacking velocity and acceleration of the whole limb. Thus, the observed small finger force when generating and adjusting a *pp* tone could be a consequence of a small mass with a low stiffness depressing the key at a low velocity and acceleration, while the large force at *ff* was due to the combined effect of a large mass, high joint stiffness, and a high attacking velocity and acceleration. These findings are in agreement with suggestions from theoretical papers, e.g., Parncutt and Troup (2002).

The present study also demonstrated that a large portion of the finger force was generated before the key reached the bottom of its displacement. This is reasonable because the hammer-driving structure commonly starts to move when the key end descends beyond the halfway point (about 6 mm in the present experimental piano) of its movement range (10.5 mm). This movement range was about 1 mm larger than the grand piano reported in a previous study (Askenfelt and Jansson, 1991). Therefore, in theory, force application can be completely terminated before the key reaches the bottom of its movement range since any force applied to the key after the event of the hammer thrust cannot contribute to sound generation. The results of the present study indeed indicated that the impulse before the moment of maximum key displacement commonly exceeded 80% of the total impulse generated. This finding disagrees with that of Harding *et al.* (1989), who have shown that finger-force profiles in

Kinoshita *et al.*: Sound pressure and keystroke force

staccato strikes have a larger peak and impulse during the period after the maximum key displacement (also the "bottoming-out" in their definition) moment than prior to it. This discrepancy appears to have resulted from the method used to determine the moment of maximum key displacement. Harding *et al.* used force data for this purpose under the assumption that the maximum finger force occurred at the moment of maximum key displacement when the key collided with the felt of the front rail. Our simultaneous recording of finger force and displacement indicates that this assumption is incorrect. As was shown in Fig. 3, the maximum force commonly appeared earlier than the moment of maximum key displacement. An extremely early case involves the trials at low SPLs, where the finger-key contact had been terminated before key displacement reached its maximum point (see "*pp*" in Fig. 3 as an example). In this case, the maximum displacement was only 6–7 mm, and therefore there was no collision of the key with the felt cushion on the front rail. At higher SPLs, on the other hand, the key vigorously hit the felt, and the time lag of the maximum force and maximum key displacement was shortened (see "*ff*" in Fig. 3). Nevertheless, the maximum key displacement was always detected after the event of maximum force, possibly due to depression of the felt which occurred after the event of key-felt collision.

The above findings also indicate that some portion of the post key-bottom force (<20% of the total impulse) is difficult to avoid, especially when producing a louder sound. There are several reasons for this. One is that when producing a louder sound, switching of the movement direction from downward to upward at an exact moment becomes more difficult due to momentum (larger segmental mass moving at a faster velocity). Under such conditions, the probability of a delayed switching of movement becomes higher. Related to the difficulty of minimizing the post-bottom finger force as well as the time of force application at a high SPL, Parlitz *et al.* (1998) provided evidence that its impulse was larger, and the time was longer for less skilled players than skilled professional players. Second, when producing a louder sound, key depression is commonly made using a coupled flexion movement at the wrist and shoulder joints (Furuya *et al.*, 2006). This maneuver facilitates a thrusting forward motion of the whole arm to accomplish a more vigorous key depression. However, it also demands a longer finger-key contact time for longer force application to the key.

In the present study, the on-off timing of the key contacting with the front rail was not monitored. This was because in many of the keystrokes at *pp* and *p* tones and even some at an *mf* tone, the key did not hit the felt (bottom) of the front rail with the present fast *staccato* touch. According to Askenfelt and Jansson (1990), however, the key bottom contact is an important mechanical event providing sensory feedback from the instrument, decisive for the player's ability to perform the desired timing and synchronization of the notes. Therefore, in a future study, analysis of finger force should include this temporal information for the better understanding of finger-force control by pianists (Parlitz *et al.*, 1998).

B. Inter-subject variation in finger force

A relatively large inter-pianist variation was noted in the maximum force and impulse during key attack, showing that the same target SPL could be attained by the application of different levels of finger force. This may be of interest from the view of stress reduction since the repeated application of excessive finger force has been considered a risk factor of over-use injuries in musicians including pianists (Amadio and Russotti, 1990; Caldron *et al.*, 1986; Fry, 1991; Zaza and Farewell, 1997). The present study also demonstrated that the inter-pianist differences in force were related to the mass involved in key depression (effective mass) but not to an individual's body weight. The force was also unrelated to key acceleration. These findings suggest that the adjustment of effective mass, and thus the modulation of rigidity of the muscles in the hand and arm, is a key factor for the stress-related problems induced by finger force. This may have an important implication in the prevention of over-use injuries.

C. Finger force of struck and pressed touches

The type of key touch clearly differentiated the pattern of finger-force development. For the struck touch, the initial force increase was rapid and formed a distinct key-touch-associated peak, followed by many peaks in the subsequent phase. For the pressed touch, on the other hand, the initial force increase was less steep without any initial peak, and it showed less force fluctuation in the mid-depression phase than the struck touch. Surprisingly, despite these clear differences in the force profile, we found that the maximum force and impulse exerted on the key were quite similar between the two touch types at the same SPL except when the subjects exerted a large force at *ff*. These findings, therefore, suggest that, at least with *staccato* articulation, SPL is largely determined by how forcefully the pianists press or strike the key, and not by how smoothly they apply force.

Goebl *et al.* (2005), who recently compared these two types of keys touch, reported that with the pressed touch, both of the key and hammer velocities developed much smoother and their spatio-temporal features had better correspondence than those with the struck touch especially when producing louder sound. Indeed, the velocity profiles shown in their study clearly described that for the struck touch, there was a period immediately after the finger-key contact where key velocity was increased markedly, but hammer velocity remained nearly unchanged. This observation led them to suggest that the pressed touch was more efficient than the struck touch. We, on the other hand, found that at the same SPL of *ff*, the pianists exerted a significantly larger finger force with the pressed touch than the struck touch. Therefore, contrary to the suggestion by Goebl *et al.*, our data reflecting the effort of force to move the key suggest that the pressed-touch method is less efficient (possibly demanding greater muscular effort) than the struck-touch method in loud sound production.

The computed descending acceleration of the key was significantly less with the pressed touch, possibly due to an inability to promote a rapid acceleration of the key from zero acceleration within its limited movement range. The results

indicated that compensation for this problem occurred by increasing the mass involved in pressing the key, which was indeed overcompensated. The reason for this overcompensation was uncertain, but we speculate that the adjustment of effective mass, which involved both the number of muscles and limb stiffness, may be more difficult than the adjustment of limb acceleration.

D. Finger noise

In agreement with the findings of previous studies (Askenfelt, 1994; Goebl *et al.*, 2004; Koornhof and van der Walt, 1994), in the present study, a short period of finger noise prior to a large piano sound was exclusively detected in the vast majority of strokes with the struck touch. In the pressed touch, there was also a finger noise-like single prior to the piano sound in some trials above *mf* levels. This smaller number of observations in the pressed touch is obviously due to the fact that the finger noise is caused mainly by the fingertip hitting the key. Detection of a similar noise without hitting the key in the pressed touch indicated that the noise could also be caused by the fingertip or the nail scraping the key surface when depressing the key at fast speed.

The present study also demonstrated that the amplitude of the initial peak force appearing within the first 5 ms contact period was related closely to the SPL of touch noise. The finding that a descending movement of the key did not commonly occur until the end of this short force pulse suggested that a large portion of kinetic energy resulting from this initial force application had been dissipated as noise rather than key movement. An additional observation related to touch noise was that, in all subjects, the SPL exceeded 90 dB at *ff*, which happened to be equivalent to the SPL of piano sound at *pp*. The noise must therefore be at an audible level even for listeners at a certain distance from the piano. Goebl *et al.* (2004), who studied the perception of touch noise in tone production, reported that only some trained musicians were able to distinguish between a struck and a pressed touch using this noise as a cue. When the listeners were unable to hear the touch differences, they tended to rate louder tones as being struck and soft tones as being pressed. The authors therefore speculated that the pure aural effect of touch noise on piano sound was relatively small. The finding in this study that the SPL difference between piano sound and finger noise was always greater than 15 dB supports their assumption; although the touch noise is audible, the subsequent piano sound is large enough as well as short enough (<50 ms) to mask its effect, due to the physiological mechanisms of backward masking (Yost, 2000).

E. Limitations of the study

Only one kind of small upright piano was used in the present study. In the previous studies of finger-force measurement, on the other hand, a grand piano has been commonly used (Askenfelt, 1994; Askenfelt and Jansson, 1991; Harding *et al.*, 1989; Parlitz *et al.*, 1998). There are apparent differences in the mechanics and acoustics between upright and grand pianos (Fletcher and Rossing, 1991, 1998). The grand piano has a more complex key action mechanism with a greater number of moving parts (i.e, a larger mass and more mechanical frictions in general) to move the hammer vertically (i.e., a greater effect of gravity) compared with the upright piano having a less complex mechanism with a horizontal hammer motion. These mechanical differences may allow pianists with a better feel of the hammer weight, a more detailed and accurate regulation of key action, and thus a finer sound voicing for the grand piano. Pfeiffer (1978) earlier compared mechanical efficiency of grand and upright piano actions by dropping weights onto piano keys and determining the resulting kinetic energy imparted to the hammers (see summary tables (pp. 314–315) given in a book by Fletcher and Rossing, 1991). The upright piano has a slightly higher mechanical efficiency of this energy transfer than the grand piano at below an *mp* tone, which is reversed from an *mf* tone on. Therefore, in the upright pianos, there could have been less finger force required at very soft dynamics to produce a piano tone while greater finger force at louder dynamics compared with the grand piano.

Sound dampers are also more effective in grand than upright pianos, providing a difference in the quality of the tone produced. Acoustical advantages of the grand piano due to its larger size with longer strings and horizontal positioning of the soundboard also provide a greater range of SPL due to larger resonance and more sound reflection, and more radiated sound than the upright piano. All of these features potentially influence the finger force action on the keys, and thus the force-sound relationship. Some of the differences between our results and those of previous studies as discussed above, therefore, can be attributed to these upright and grand piano differences. A comparative study using other kinds of upright as well as grand pianos is therefore needed in the future.

Another limitation of this study was that we measured only the vertical component of finger force under the assumption that forces acting in the fore-aft and medial-lateral directions were small. The validity of this assumption may need to be examined by the use of a three-dimensional force sensor under the experimental conditions similar to the present study.

Amadio, P. C., and Russotti, G. M. (**1990**). "Evaluation and treatment of hand and wrist disorders in musicians," Hand Clin. **6**, 405–416.

Askenfelt, A. (**1994**). "Observations on the transient components of the piano tone," *Proceedings of the Stockholm Music Acoustic Conference*, Royal Academy of Music, Stockholm, **79**, pp. 297–301.

Askenfelt, A., and Jansson, E. V. (**1990**). "From touch to string vibrations. I: Timing in the grand piano action," J. Acoust. Soc. Am. **88**, 52–63.

Askenfelt, A., and Jansson, E. V. (**1991**). "From touch to string vibrations. II: The motion of the key and hammer," J. Acoust. Soc. Am. **90**, 2383–2393.

Askenfelt, A., and Jansson, E. V. (**1992**). "On vibration sensation and finger touch in stringed instrument playing," Music Percept. **9**, 331–350.

Caldron, P. H., Calabrese, L. H., Clough, J. D., Lederman, R. J., Williams, F., and Leatherman, J. (**1986**). "A survey of musculoskeletal problems encountered in high-level musicians," Med. Probl. Perform. Art **1**, 136–139.

Fletcher, N. H., and Rossing, T. D. (**1991**). *The Physics of Musical Instruments* (Springer–Verlag, New York).

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments* (Springer–Verlag, New York).

Furuya, S., Aoki, T., and Kinoshita, H. (**2006**). "Control of upper extremity movements in expert pianists when striking the piano keys at various combinations of sound volume and striking tempo," J. Soc. Biomecha-

nisms **30**, 151–155 (in Japanese with English abstract).

Fry, H. J. H. (**1991**). "The effect of overuse on the musician's technique: A comparative and historical review," Int. J. Arts Med. **1**, 46–55.

Goebl, W., Bresin, R., and Galembo, A. (**2005**). "Touch and temporal behavior of grand piano actions," J. Acoust. Soc. Am. **118**, 1154–1165.

Goebl, W., Bresin, R., and Galembo, A. (**2004**). "Once again: The perception of piano touch and tone. Can touch audibly change piano sound independently of intensity?" *Proceedings of the International Symposium on Musical Acoustics*, The Acoustical Society of Japan, pp. 332–335.

Harding, D. C., Brandt, K. D., and Hillberry, B. M. (**1989**). "Minimization of finger joint forces and tendon tensions in pianists," Med. Probl. Perform Art **4**, 103–104.

Keppel, G. (**1991**). "Analytical comparisons among treatment means," in *Design and Analysis, A Researcher's Handbook*, Prentice–Hall, Upper Saddle River, NJ, pp. 109–184.

Kinoshita, H., Kawai, S., and Ikuta, K. (**1995**). "Contributions and coordination of individual fingers in multiple finger prehension," Ergonomics **38**, 1212–1230.

Koornhof, G. W., and van der Walt, A. J. (**1994**). "The influence of touch on

piano sound," *Proceedings of the Stockholm Music Acoustic Conference*, Royal Academy of Music, Stockholm, **79**, pp. 302–308.

Palmer, C., and Brown, J. C. (**1991**). "Investigations in the amplitude of sounded piano tones," J. Acoust. Soc. Am. **90**, 60–66.

Parlitz, D., Peschel, T., and Altenmuller, E. (**1998**). "Assessment of dynamic finger forces in pianists: Effects of training and expertise," J. Biomech. **31**, 1063–1067.

Parncutt, R., and Troup, M. (**2002**). "Piano," In *The Science and Psychology of Music Performance. Creative Strategies for Teaching and Learning*, edited by P. Parncutt and G. Mcpherson (University Press, Oxford), pp. 285–302.

Pfeiffer, W. (**1978**). *The Piano Hammer: A Detailed Investigation Into an Important Facet of Piano Manufacturing* (Verlag Das Musikinstrument, Frankfurt).

Yost, W. A. (**2000**). *Fundamentals of Hearing* (Academic, San Diego).

Zaza, C., and Farewell, V. T. (**1997**). "Musicians' playing-related musculoskeletal disorders: An examination of risk factors," Am. J. Ind. Med. **32**, 292–300.

Vibrotactile transduction and transducers

Bruce J. P. Mortimer,[a] Gary A. Zets,[b] and Roger W. Cholewiak[c]
Engineering Acoustics, 933 Lewis Drive, Suite C, Winter Park, Florida 32789

The body's sense of touch is potentially a versatile channel for the conveyance of directional, spatial, command, and timing information. Most practical implementations of vibrotactile systems require compact, light-weight actuators that can be mounted against the body. Eccentric mass motors are widely used for this application, yet their output is limited and the effects of loading on the transducers due to the skin and mounting arrangement have been largely ignored. Conventional linear actuators are well suited as vibrotactile transducers and can provide high output, but are typically limited to laboratory research due to their large size and cost. The effect of loading on various practical vibrotactile transducers is investigated using a skin impedance phantom and measuring the transducer displacement with respect to additional mass loading. Depending on the transducer design, loading can dramatically reduce the vibratory displacement and, in the case of eccentric mass motors, also increase the operating frequency. In contrast, a new linear actuator design can be designed to be almost independent of skin loading, by considering the mechanical impedance of the load and optimizing the transducer contact area. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2715669]

I. INTRODUCTION

The human body relies on multiple receptors in the skin that, when combined, provide a wide range of information to the brain regarding objects that come close to, or touch the body. These receptors are sensitive to different stimuli including temperature, pain, pressure, and vibration. The senses of touch are linked with the neuro-motor channel, which makes it uniquely tied to movement. It is also postulated that automatic motor responses to touch may occur more quickly than with the other senses.

Multichannel vibrotactile displays have been used to present information to individuals regarding their posture (Wall *et al.*, 2002), position, or location in three-dimensional space and to enhance spatial awareness (Rupert *et al.*, 1994; McGrath *et al.*, 2004; Cholewiak and Beede, 2005; and van Veen and van Erp, 2003). Tactile displays have also proven to be effective communication channels, conveying information to wearers by appropriate sequences and patterns (Brewster and Brown, 2004; Terrence *et al.*, 2005). Low level vibrotactile signals have also been shown by Collins (Richardson *et al.*, 1998; Priplata *et al.*, 2006) to be effective in the treatment of peripheral neuropathy.

Transducers used in vibrotactile displays have been both variable in approach and performance. The actual requirements for transducers depend on the intended application and tactile interface requirements. The body's response to tactile stimuli is somewhat complex depending on stimulus characteristics, body location, transducer geometry, and a large number of psychophysical factors. This paper first reviews studies on the human sensory response to tactile stimuli and relates these to requirements for vibrotactile displays. New and existing transducer actuator designs are assessed for their application in wearable vibrotactile transducers.

II. RESPONSE OF HUMAN SKIN TO VIBRATION

The sensitivity of the body's skin receptors to vibratory displacement is well known (Greenspan and Bolanowski, 1996; Cholewiak and Collins, 1991). Four receptor systems are thought to contribute to detection of vibrotactile stimuli at threshold under normal conditions. In glabrous (smooth) skin, Pacinian corpuscles (PCs), Meissner's corpuscles, Merkel's disks, and Ruffini cylinders are the structures associated with the four psychophysically defined channels (Bolanowski *et al.*, 1988). Of these, the PCs are the most sensitive. At 250 Hz, the sensitivity of human glabrous skin to displacement is less than 1 μm. A similar set of channels have been proposed for hairy skin (Bolanowski *et al.*, 1994) although the delineation is less precise. For example, rapidly adapting (RA) PC-like responses are obtained from hair follicles and field receptors (Greenspan and Bolanowski, 1996, pp. 42–43). Furthermore, the overall sensitivity is less: that of the PC-like channel is 20 dB less than in glabrous skin (about 1/10 as sensitive).

Mechanotransduction is the process by which mechanical displacement is converted into neural action potentials. The receptor system has to cover a broad range of mechanical sensations—therefore requiring the skin to made up of a multitude of receptors, each with different sensitivity and frequency dependence. The density distribution of receptors also varies across the skin, being more concentrated in body regions involved in exploration and manipulation. Receptors also have different rates of adaptation to stimuli. PC receptors, for example, are RA in that their sensitivity to a continuous stimulus decreases with time.

The mechanotransduction process can be characterized

---

[a] Electronic mail: bmort@eaiinfo.com
[b] Electronic mail: zets@eaiinfo.com
[c] Electronic mail: rcholewi@princeton.edu
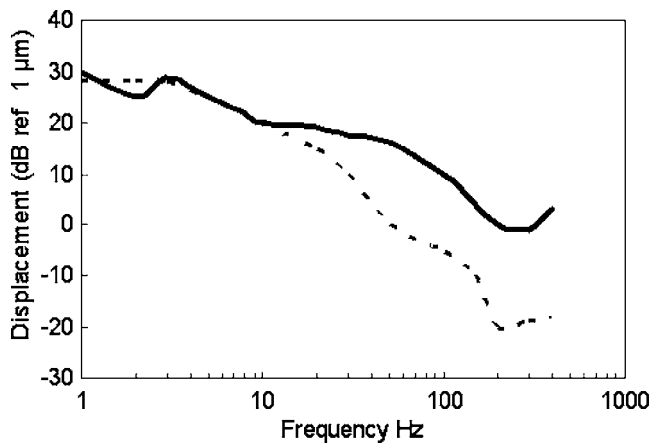
FIG. 1. Threshold of vibration detection vs frequency in smooth, or glabrous, skin (dashed line, adapted from Bolanowski *et al.*, 1988) and curve for hairy skin (solid line, adapted from Bolanowski *et al.*, 1994). Hairy skin is about 20 dB less sensitive in the high-frequency range than the glabrous skin curve. It also has a slightly different shape, especially in the <100 Hz range.

as a *system* in terms of *sensitivity* and *resolution* (both spatial and temporal). Sensitivity is dependent on the location of the stimulus on the surface of the body, the sex and age of the subject, temperature, the frequency and the wave form of the stimulus, the area of the contactor, the duration of the stimulus, and the presence of a surround. Hairy skin, like that of glabrous skin, is more sensitive in the range 200–300 Hz (region where PC-like receptors are operational) and is shown in Fig. 1. For this reason, vibrotactile transducers have mostly been implemented over this range of frequencies.

Successiveness limen relates to temporal resolution— How closely spaced a series of stimuli can be for a person to distinguish them as separate. The limit of human perception of temporal aspects of tactile stimulation is about 20 ms. This is considerably more limited than that for acoustic stimuli, but this factor alone does not limit the data transfer rate via tactile channels.

Spatial resolution can be determined in a number of ways. Two common methods are either to measure the two point limen, a test that determines the discrimination threshold for recognizing two closely spaced sensations or a localization test which measures the accuracy with which the location of a touch can be identified. Spatial resolution depends on the location of the body being studied and the type of stimulus (Cholewiak *et al.*, 2004; Tan *et al.*, 2003; van Erp and Werkhoven, 1999; Sherrick *et al.* 1990).

Other relevant variables in determining vibrotactile spatial acuity can be found in studies in which two (or more) stimuli were used. Profound interactions are revealed among stimuli, particularly when presented close together in space and in time (e.g., Soto-Faraco *et al.*, 2004; Weisenberger, 1994). For example "masking" interactions that occur when two vibratory stimuli are presented close together in space and/or time, finding that the closer the two are to one another (on either dimension), the more the perceived test stimulus sensation magnitude is reduced. If noise is band limited, the resultant masking effect may be restricted to a receptor band;

a stimulus within the channel is affected by the masking while stimuli in other bands can remain unaffected. These effects have been used to prove the channel independence in the skins sensing system.

In contrast, spatial summation is an effect whereby there is a reduction in detection threshold as a function of the contact area (e.g., Verrillo, 1963). This effect only occurs in the Pacinian system. Such a mechanism has been explained as the addition of energy from larger and larger areas of stimulation. Force was not controlled in these studies, only initial penetration. Craig and Sherrick (1969) subsequently demonstrated that holding force, pressure, or indentation constant as contactor area is increased produced different spatial summation results especially at higher frequencies. The static indentation of a vibrating contactor even over a range of 1 mm or less has a significant effect on vibrotactile thresholds of up to 14 dB (Makous *et al.*, 1996). This effect is partly due to the change in skin impedance (Fritschi *et al.*, 2004).

A rigid surround around a moving contactor offers a method for damping out surface waves from the vibrating stimulus (Verrillo, 1962; Gescheider *et al.*, 1978; Cholewiak and Collins, 1991; Goble *et al.*, 1996). The effect of the surround may increase or decrease vibrotactile threshold depending on the nature of the underlying receptor systems and also depends on contactor size and the radial gap between contactor and surround (Cholewiak and Collins, 1991; Greenspan and Bolanowski, 1996).

The subtle interaction of experimental configurations with outcome illustrates the difficulties found in psychophysical evaluation of vibrotactile phenomena. As we will show in our analysis of vibrotactile transducer designs, there are distinctly different operational modes for transducers. Experimental conditions such as frequency, stimulus, and mounting are all known to affect the transducer operation while the training, age, fatigue level, and transducer placement are human factors that must be carefully analyzed. In previous work on the forearm and hand these factors have been relatively well managed in precise laboratory based experiments. However, we will show that care must be taken when extending the findings from these experiments to wearable tactile arrays because the results will now be found to depend on transducer behavior and mounting.

III. VIBROTACTILE TRANSDUCER DESIGNS

Fundamental to any transducer operation is the balancing of forces—Those due to the actuator and the reaction forces that hold it in equilibrium (often from the skin load and/or mounting configuration). The inertial reference is the "ground" or zero from which these forces operate.

In wearable vibrotactile systems, the ground is reached through the mechanical impedance of the body. In laboratory systems, the ground reference may be external to the body, for example directly through a rigid and massive frame. This is shown in Fig. 2.

Directly ground referenced vibrotactile transducer designs include diagnostic devices described as vibrometers by Dellon (1983). These devices produce and monitor the me-
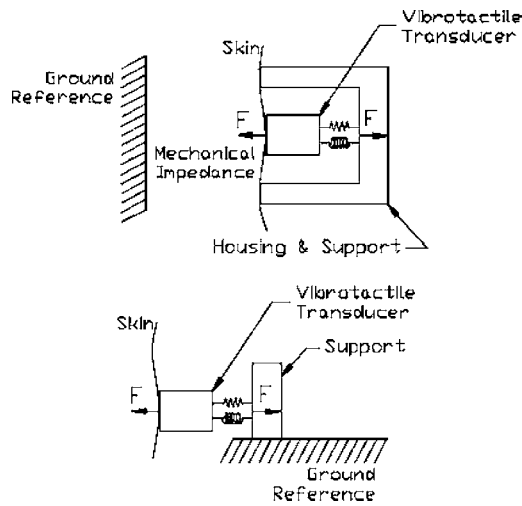
FIG. 2. Schematic showing a definition of force balances and an inertial reference; (1) wearable vibrotactile transducer with an inertial reference "ground" through the impedance of the skin (2) externally referenced vibrotactile transducer.

FIG. 3. Magnitude of skin load impedance at 100 (long dash) and 250 Hz (solid) vs diameter of contactor in centimeters. At low frequencies the real part of the skin impedance is not frequency dependent and is shown with dotted lines.

chanical stimulation against the skin using a moving mass contactor and are primarily used to determine sensitivity thresholds. The housing of the device is large and mounted to an external rigid stand and support (the ground reference), and only the moving mass makes contact with the skin. Some recent haptic finger displays are also ground referenced (Pasquero and Hayward, 2003).

Extending transducer design concepts to wearable vibrotactile displays introduces stringent size, weight, and efficiency requirements. A somewhat overlooked requirement is that the vibrotactile transducer should also produce a displacement output that is relatively independent of loading. This implies that one must account for the inertial forces using a ground reference system that is linked via the mechanical impedance of the body. The transducer model will also have to consider the additional mechanical impedance elements contributed by the mounting arrangement that holds the transducer against the body.

A. The skin load

Numerical values for the skin mechanical impedance components can be found in Franke (1951) and Moore and Mundie (1972). These references show that skin tissue has the mechanical input impedance of a fluid-like inertial mass, a spring-like restoring force, and a viscous frictional resistance—the skin is, therefore, a viscoelastic medium. The numerical magnitude of each component in the skin impedance depends on the frequency and contact area. This effect is shown in Fig. 3.

The magnitude of the skin impedance is shown for two frequencies. For comparison the magnitude of the resistive part is also shown. As can be expected, the resistive loading of the skin is shown to increase with increasing contactor diameter.

Investigations into the mechanical impedance of the skin at the finger have shown a nonlinear increase in stiffness with contactor indentation until a maximum indentation of approximately 3 mm, constrained by the compression of the tissue between contact surface and bone of the finger (Fritschi et al., 2004). This nonlinearity is also observed for other areas of the body (Franke, 1951).
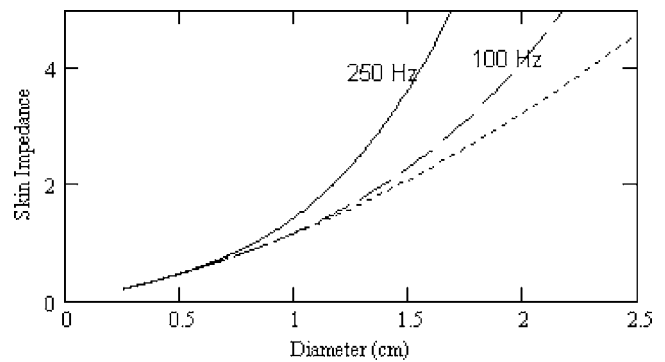
B. Experimental setup and rationale

The mechanical impedance characteristics of the skin can be simulated with a skin model (for example, Royston et al., 1999; Hall et al., 1997). In our experiments we constructed a layered viscoelastic phantom consisting of a gelatin gel base mixture (composition 50 g gelatin/l water), ≈5-mm-thick agar gel top coupling layer (composition 16 g agar/l water) and a membrane surface (0.32 mm latex rubber sheet). The phantom was constructed in a 110×130 mm, 25 mm deep aluminum box, with a rubber membrane pulled over the front surface and secured with two o-rings. The measured stiffness of this setup was 4.9×10^5 dynes/cm (measured with a 9.6-mm-diam contactor) which is comparable to that of hairy skin. Transducers could be mounted to this phantom using 0.03-mm-thick double sided tape (3M). Displacement was measured using an analog fiberoptic sensor (Philtec model D63).

The effect of mounting on the operation of wearable vibrotactile transducers has been somewhat ignored in previous studies. Mounting may add mass, compliance, damping or depress the transducer against the skin load. Depending on transducer design these mounting parameters can have a profound effect on the transducer's operating characteristics. In the worst case, the moving element can even become immobilized, with the driving energy being dissipated as heat rather than vibration.

C. Categories of vibrotactile transducers

Vibrotactile transducers fall into two broad categories depending on their design and inertial reference. The first category of vibrotactile transducers is that of inertial shakers. A typical example would be an eccentric mass motor or pager motor. The second category is that of a linear actuator which drives a contactor or moving element against the skin.

Inertial shaker designs utilize the motion of an internal mechanism to translate force to the transducer housing and consequently the adjacent structure. An example is the ubiquitous pager motor. Shaker designs usually use an electro-

2972   J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Mortimer et al.: Vibrotactile transduction and transducers
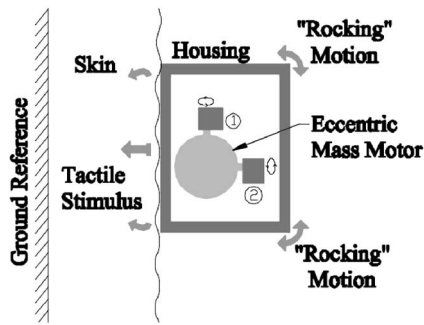
FIG. 4. Pager motor operation. Two configurations are shown: (1) Eccentric mass rotating orthogonal to the skin and (2) pancake motor with an eccentric mass rotating on an axis parallel to the skin.

magnetic drive configuration and may in some designs include a spring to introduce resonance which limits the operating frequency but may increase overall efficiency (Cholewiak and Wollowitz, 1992).

Actuator designs are usually characterized by linear plunger vibration. Designs have included electromagnetic, moving coil, variable reluctance, piezoelectric, and bimorph benders as drive mechanisms. These often operate at resonance and are, therefore, prone to damping and are often variable in their operation. In a laboratory environment researchers have often controlled their experiments by measuring some form of contactor displacement and applying feedback to keep the output at constant displacement.

IV. ECCENTRIC MASS TRANSDUCERS

A. Device description

A pager motor is usually a dc motor with an eccentric mass load such as half-circular cylindrical weight that is mounted onto the motor's shaft. From the conservation of angular momentum, the eccentric mass imparts momentum to the motor shaft and consequently the motor housing. The angular momentum imparted to the motor housing will depend on the mounting of the motor housing, the total mass of the motor, the mass of the eccentric rotating mass, the radius of the center of mass from the shaft, and the rotational velocity. In steady state, the angular momentum imparted to the housing will result in three-dimensional motion and a complex orbit. This implementation applies forces in a continually changing direction confined to a plane of rotation of the mass. Thus the resultant motion of the motor housing is three dimensional and complex.

The operation of a pager motor vibrotactile transducer is shown schematically in Fig. 4. Two orientations of eccentric mass are possible, either a cylindrical motor with an eccentric mass rotating orthogonal to the skin (1) or a pancake motor with an eccentric mass rotating on an axis parallel to the skin (2). It is common to enclose the pager motor within a housing. The housing protects the orbit of the eccentric mass, avoiding jams and associated high motor fault current. The housing vibrates and makes up the contactor against the skin. As we have described, the motion of the adjacent housing will not be rectilinear. Further, all of the housing surfaces vibrate.

There are a number of important effects associated with the mounting of pager motors as wearable vibrotactile transducers. The complete housing surface makes up the vibrating contactor, thus the motion that is translated to an adjacent body will depend on the loading on the motor housing—from the conservation of momentum, the greater the mass loading on the motor (or transducer housing) the lower the vibratory velocity and the perceived amplitude of the stimulus. Any change in the mass loading on the motor will also change the torque on the motor. If the motor is a dc motor driven with a constant voltage, the angular rotation rate and thus frequency of vibration will also change. Thus depending on the mounting configuration, the displacement into the skin and perception of a vibratory stimulus is variable in frequency and level. In fact it is not possible to simultaneously and independently control output vibration level and frequency. This is obviously undesirable from a control standpoint, and in the limiting case, a highly loaded transducer would produce minimal displacement output and thus be ineffective as a tactile stimulus.

The second effect associated with the operation of pager motors as vibrotactile transducers is that the vibration of the motor or housing against the skin is complex—There is no uniform velocity on any surface of the contact area.

B. Experiments

The performance of three different pager motor transducers under various loading conditions was investigated. The first two were our own designs of a pager motor mounted in a 17.9-mm-diameter aluminum can; the units were 11 and 9.2 g, respectively. A third pager motor design (Piateski, 2005) uses a pancake eccentric mass motor mounted in a rigid, lightweight plastic material—The pancake motor was sealed with glue, and then molded in a plastic block 18.4 mm long, 17 mm wide, and 6 mm thick. Each of the pager motor designs was attached to the top surface of a skin model (described in Section III B) using thin double sided tape (3M). Displacement and frequency were measured using a probe mounted directly above the pager motor. Loading on the transducer was varied by the addition of mass elements symmetrically onto the back surface of the designs. Figure 5 illustrates the effect of loading on pager motor performance.

As may be expected, the addition of a mass load on any of the pager motor transducer designs causes a decrease in vibratory displacement and an increase in frequency. The pancake pager motor (square markers) shows the highest initial unloaded displacement due to its low mass, however, it is quickly damped with the addition of mass loading. In the unloaded condition, the mass and contact area of the transducer can be designed to maximize the resonant characteristics of the skin's mechanical impedance. However, in all of the designs studied, these gains are quickly eroded by even small amounts of loading as may be expected in any practical wearable transducer setup.
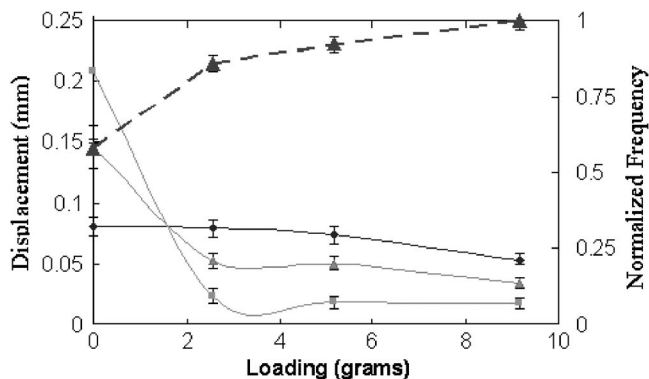
FIG. 5. Displacement (millimeter peak) output of three different pager motor designs vs mass loading (grams) measured on a skin phantom are shown in the lower three solid line curves (the diamond markers show the response for the 11 g, triangles the 9.2 g can designs, and square markers the pancake motor design). The effect of loading on a pager motor mounted against the skin (upper curve, dashed line) shows an increase in frequency and decrease in displacement with increasing loading force (measured for the pancake motor design). In all cases, the drive voltage was held constant at 3.1 V.

V. LINEAR ACTUATOR VIBROTACTILE TRANSDUCERS

A. Device description

Linear actuators such as the Bruel & Kjaer 4810 minishaker have been used in the study of tactile perception (for example, Goble *et al.*, 1996). When these transducers are combined with a measurement of acceleration (or displacement) they can serve as effective laboratory systems. Extending the use of these actuators to wearable applications, however, has not been possible because they typically are ground referenced, damped by the skin loading, or have a high mass. The B&K 4810, for example, weighs 2.4 kg and is 7.5 cm tall and wide.

Figure 6 shows in schematic form, a wearable alternative to such massive (but precise) linear actuator shakers. In this design, the contactor is the predominant moving mass, driving the skin with perpendicular movement that is independent of the loading on the housing (over a normal range of forces used in wearable designs).

B. Experiments

The effect of mass loading on the housing is shown in Fig. 7. The wearable linear actuator housing was mounted onto the previously described skin model using double sided tape with the contactor free to vibrate against the phantom.
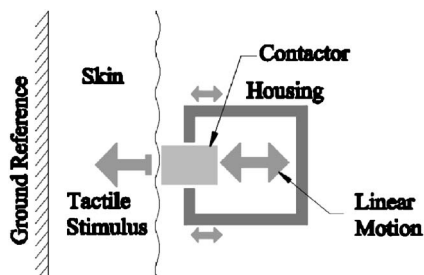
FIG. 6. Wearable linear actuator operational schematic. Both the housing and contactor are in simultaneous contact with the skin load but the contactor is designed for maximum displacement.
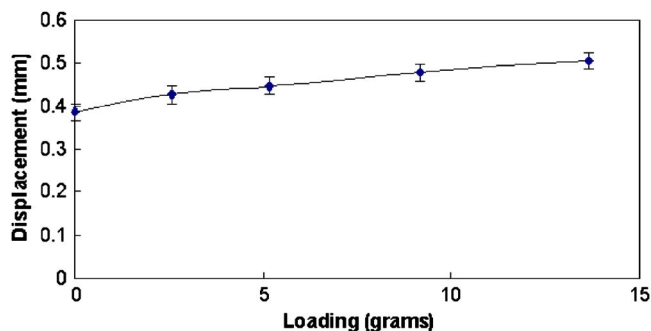
FIG. 7. (Color online) Wearable linear actuator contactor displacement (mm peak) vs various mass loads placed on the housing to simulate the effects of *in situ* loading of a wearable transducer. This should be compared to Fig. 5, which shows the pronounced effect of mass loading on the vibratory output.

Mass loading was on the housing and the displacement was measured on the back side of the contactor. The wearable linear actuator was driven at 250 mA rms at 250 Hz (sinusoid).

Figure 7 shows an increase in contactor displacement output as mass loading is added to the wearable linear actuator housing. This is in contrast to the results for inertial shaker transducers shown in Fig. 5, where displacement output dropped dramatically, and shows the role of the housing in the wearable linear actuator design.

C. Wearable transducer design considerations

The housing serves as a reaction mass for the motion of the contactor. In wearable applications, the combined weight of the housing and contactor should be minimized. However, the area of the housing face that is in contact with the skin load introduces a reactance due to the mechanical impedance of the skin. This mechanical impedance is predominantly a mass load over the operating range of the wearable linear actuator. Thus the actual housing mass and the component of the skin mechanical impedance that is in contact with the housing make up the total lumped housing mechanical impedance.

The effect of the load and the housing mass can be explained in the "free-body diagram" shown in Fig. 8. The linear motor in the electromagnetic wearable linear actuator (tactor) generates two equal and oppositely directed forces, one on the contactor-reciprocating magnet assembly, and the other on the tactor housing which contains the stator.

In this diagram the velocity of the housing is represented by V_H, the contactor velocity is V_C, the mass of the tactor housing which contains the motor stator is M_H, the suspension spring mechanical compliance is C_S, and the contactor-moving mass is M_C. The skin component load on the contactor face comprises mass M_S, mechanical compliance C_S, and mechanical resistance R_S. Similarly, the skin component load on the housing comprises M_{SH}, R_{SH}, and C_{SH}.

The relationship between the transducer contactor area (and consequently the mechanical impedance M_C), contactor displacement, and the actual perception of vibratory stimulus needs some further explanation. Section II described the limitations of cutaneous sensitivity and the effect of contactor area, transducer surround and body locale on the thresh-
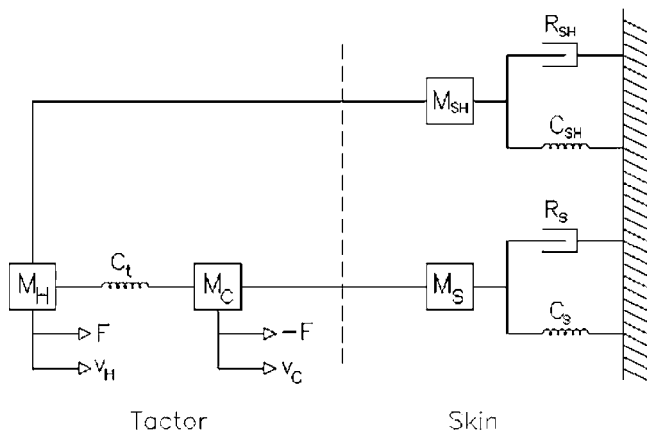
FIG. 8. Free-body diagram showing the operation of the wearable linear actuator. The skin load is acted on by a moving contactor with mass M_C as well as the housing M_H. In each case the skins' mechanical impedances are different.

FIG. 9. Calculated normalized skin stimulus (contactor area × relative contactor displacement at 100 and 250 Hz) vs diameter of contactor in centimeters using the skin's mechanical impedance characteristics, as found in Franke (1951).

old of vibratory sensation. The initial depression of the contactor, the applied force, and contactor pressure can all be considered as experimental configuration variables in human psychometric testing. Craig and Sherrick (1969) showed that these variables can all play a role in the perception of vibration stimuli and must be carefully controlled in comparative studies. A study by Lamoré and Keemink (1988) also included the effect of surround, body location, and force as variables in determining the absolute perception of vibration stimuli. Changing these variables can be partially anticipated from the physics of the body tissue (Franke, 1951) and explained as a change in skin stiffness.

From a transducer design perspective we can simplify the analysis by concentrating on the role of the actuator in the mechanotransduction process. For a fixed frequency (and geometrical configuration), the mechanical impedance of the skin presented to the contactor depends only on the contactor diameter, as shown in Fig. 3. If contactor displacement was held constant, more power would be delivered to the skin load with increasing contactor diameter and, depending on the initial static contactor force, we would anticipate an increase in perceived vibrational stimulus and at higher frequencies spatial summation. However, transducers are usually not constant displacement devices but are rather limited by their maximum force generation capability and the mechanical acoustic load that they are acting against.

If we keep the transducer drive force (F shown in Fig. 8, not to be confused with static force) constant, we can compare the displacement "performance" of the transducer for various contactor diameters. From the contactor area we can compute the total mechanical impedance presented to the contactor and also the contactor relative displacement into this load (from the admittance). If we define the skin stimulus to be the result of a combination of the contactor area x relative contactor displacement, Fig. 9 shows a plot of skin stimulus calculated at 100 and 250 Hz for various contactor diameters. In this graph we have normalized the skin stimulus to the maximum stimulus for each frequency trace.

Figure 9 clearly describes a range of contactor diameters that will produce an optimum transducer stimulus for a con-

stant transducer force capability. Just less than 1 cm in diameter is the preferred contactor diameter for 250 Hz operation. With the contactor diameter set, and the mass of the actual elements chosen to be small, the linear actuator design can be optimized by choosing the remaining elements (spring compliance and housing impedance) for contactor resonance at 250 Hz when the transducer is mounted against the mechanical impedance of the skin.
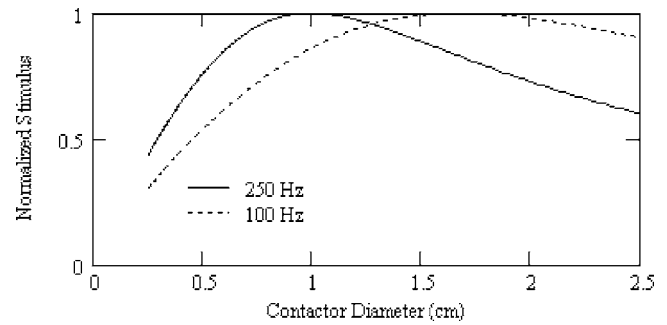
The final wearable linear actuator design is lightweight (17 g) and small enough (30 mm diameter, 8 mm deep) to provide for the array densities required in many vibrotactile display applications (e.g., McGrath, 2004). The housing serves as a surround that simultaneously contacts the skin and is separated from the moving contactor by a radial gap. The diameter of the moving element is 7 mm, centered in a 9-mm hole in the top surface of the tactor so as to provide a stationary surround with a 1-mm gap [see Gescheider et al. (1978), and Lamoré and Keemink (1988)]. The tip of the contactor protrudes 0.5 mm above the surface of the surround to ensure firm contact with the skin and this also has the effect of increasing the starting indentation force which is known to increase vibrotactile sensitivity (Craig and Sherrick, 1969; and Lamoré and Keemink, 1988).

VI. DISCUSSION

Selecting the transducer specifications for vibrotactile applications requires some careful consideration, especially in applications that require wearable transducers. The human tactile sensory response is limited in frequency response, temporal and spatial resolution. Tactile conveyance of information is more limited in bandwidth than the other senses. However, the body's ability to discern some level of frequency, amplitude, spatial, and temporal information is well known, and this information can be enhanced by multilocation tactile stimuli or spatial cuing. It follows that reasonable transducer specifications should be equivalent or better than the human receptor system's own response limitations. Reasonable technical requirements would include a frequency response to 300 Hz, a displacement output that exceeds 24 dB (to account for noise) above the threshold for sensitivity for a hairy skin body location and a rise time of less than 5 ms.

Extending vibrotactile stimulation to multiple locations requires wearable transducers that are small and have the ability to generate localized (punctuate) sensations. The design of a moving contactor with a rigid surround would appear to offer some advantages in restricting the signal to a narrow area, and as we have described earlier, a surround can also be used as a reaction mass in a compact linear actuator design.

Finally, regarding the temporal resolution of such systems, one has to consider that the start up (spin-up) times for pager motors can be relatively long, on the order of 100 ms or so. This is somewhat longer than the skin's temporal resolution, thus can limit data rates. If the vibrotactile feedback is combined with other sensory feedback such as visual or audio, the start-up delay has the potential of introducing disorientation. The disorienting effects of delayed auditory feedback for normal speech have been well-documented (Mates and Aschersleben, 2000). Although resonant, the wearable linear actuator has a rise time of less than 5 ms.

VII. CONCLUSIONS

The choice of vibrotactile transducers requires careful evaluation when they are to be used as wearable transducers. We have shown that the loading of the skin (considering the relative mechanical impedances of the tactor and tissue) as well as the implicit loading of the mounting assembly can have profound effects on the transducer vibratory output. Eccentric mass or pager motors have been widely used in wearable tactile displays and are particularly sensitive to mass loading that can affect both their stimulus frequencies and amplitudes. The slow rise-time of the motor also results in transducer latency which may be problematic. In contrast to pager motors, a new wearable linear actuator has been designed to have a housing that surrounds a moving contactor in simultaneous contact with the skin providing a reaction mass for the contactor. A simple moving mass model has been used to describe the operation of this transducer. The dimensions of the contactor and housing can be chosen to almost eliminate the effect of mass loading on the transducer. The wearable linear actuator has been designed as a resonant device, but the Q is low enough for the device to offer usable levels of performance over a very wide range of frequencies.

ACKNOWLEDGMENT

Bolanowski, S., Gescheider, G., and Verrillo, R. (**1994**). "Hairy skin: Psychophysical channels and their physiological substrates," Somatosens Mot Res. **11**, 279–290.

Bolanowski, S., Gescheider, G., Verrillo, R., and Checkosky, C. (**1988**). "Four channels mediate the mechanical aspects of touch," J. Acoust. Soc. Am. **84**, 1680–1694.

Brewster, S., and Brown, L. (**2004**). "Tactons: Structured tactile messages for non-visual information display," in *Proceedings of Australasian User Interface Conference 2004*, Dunedin, New Zealand, Australian Computer Society, pp. 15–23.

Cholewiak, R., and Beede, K. (**2005**). "The representation of space through static and dynamic tactile displays," in *Proceedings of the Virtual Reality International Conference*, Las Vegas, NV, pp. 22–27.

Cholewiak, R., Brill, J., and Schwab, A. (**2004**). "Vibrotactile localization on the abdomen: Effects of place and space," Percept. Psychophys. **66**, 970–987.

Cholewiak, R., and Collins, A. (**1991**). "Sensory and physiological bases of touch," in *The Psychology of Touch*, edited by M. A. Heller and W. Schiff (Erlbaum, Hillsdale, NJ), pp. 23–60.

Cholewiak, R., and Wollowitz, M. (**1992**). "The design of vibrotactile transducers," in *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr, London), pp. 57–82.

Craig, J., and Sherrick, C. (**1969**). "The role of skin coupling in the determination of vibrotactile spatial summation," Percept. Psychophys. **6**, 97–101.

Dellon, A. (**1983**). "The vibrometer," Plast. Reconstr. Surg. **71**, 427–431.

Franke, E. (1951). "Mechanical impedance measurements of the human body surface," Air Force Technical Rep. No. 6469, Wright-Patterson Air Force Base, Dayton, OH.

Fritschi, M., Buss, M., Drewing, K., Zopf, R., and Ernst, M. (2004). "Tactile feedback systems," in IROS 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai.

Gescheider, G., Capraro, A., Frisina, R., Hamer, R., and Verrillo, R. (**1978**). "The effects of a surround on vibrotactile thresholds," Sens Processes **2**, 99–115.

Goble, A., Collins, A., and Cholewiak, R. (**1996**). "Vibrotactile threshold in young and old observers: The effects of spatial summation and the presence of a rigid surround," J. Acoust. Soc. Am. **99**, 2256–2269.

Greenspan, J., and Bolanowski, S. (**1996**). "The psychophysics of tactile perception and its peripheral physiological basis," in *Pain and Touch*, 2nd ed., edited by L. Kruger (Academic, San Diego), pp. 25–104.

Hall, T., Bilgen, M., Insana, M., and Krouskop, T. (**1997**). "Phantom materials for elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1355–1365.

Lamoré, P., and Keemink, C. (**1988**). "Evidence for different types of mechanoreceptors from measurements of the psychophysical threshold for vibrations under different stimulation conditions," J. Acoust. Soc. Am. **83**, 2339–2351.

Makous, J., Gescheider, G., and Bolanowski, S. (**1996**). "The effects of static indentation on vibrotactile threshold," J. Acoust. Soc. Am. **99**, 3149–3153.

Mates, J., and Aschersleben, G. (**2000**). "Sensorimotor synchronization: The impact of temporally displaced auditory feedback," Acta Psychol. **104**, 29–44.

McGrath, B., Estrada, A., Braithwaite, M., Raj, A., and Rupert, A. (2004). "Tactile situation awareness system: Flight demonstration final report," Final Report USAARL Rep. No. 2004-10, U. S. Army Aeromedical Research Laboratory, Ft. Rucker, AL.

Moore, M., and Mundie, J. (**1972**). "Measurement of specific mechanical impedance of the skin," J. Acoust. Soc. Am. **52**(2), 577–584.

Pasquero, J., and Hayward, V. (**2003**). "STReSS: A practical tactile display system with one millimeter spatial resolution and 700 Hz refresh rate," in *Proceedings of Eurohaptics 2003*, Dublin, Ireland.

Piateski, E. M. (2005). "A tactile communication system for navigation," M.Sc. thesis, Mechanical Engineering, MIT, Cambridge, MA.

Priplata, A., Patritti, B., Niemi, J., Hughes, R., Gravelle, D., Lipsitz, L., Veves, A., Stein, J., Bonato, P., and Collins, J. (**2006**). "Noise-enhanced balance control in patients with diabetic neuropathy & patients with stroke," Ann. Neurol. **59**, 4–12.

Richardson, K., Imhoff, T., Grigg, P., and Collins, J. (**1998**). "Using electrical noise to enhance the ability of humans to detect subthreshold mechanical cutaneous stimuli," Chaos **8**, 599–603.

Royston, T., Mansy, H., and Sandler, R. (**1999**). "Excitation and response of surface waves on isotropic and nonisotropic viscoelastic half-spaces with application to medical diagnosis," J. Acoust. Soc. Am. **106**, 3678–3686.

Rupert, A., Guedry, F., and Reshke, M. (1994). "The use of a tactile interface to convey position and motion perceptions," in Vitrual interfaces: Research and applications, Lisbon, Portugal, 18–22 October 1993. AGARD CP-541:20-11 to 20-75. Neuilly Sur Seine, France, Advisory Group for Aerospace Research and Development.

Sherrick, C., Cholewiak, R., and Collins, A. (**1990**). "The localization of low- and high-frequency vibrotactile stimuli," J. Acoust. Soc. Am. **88**, 169–179.

Soto-Faraco, S., Ronald, A., and Spence, C. (**2004**). "Tactile selective attention and body posture: Assessing the multisensory contributions of vision and proprioception," Percept. Psychophys. **66**, 1077–1094.

Tan, H. Z., Gray, R., Young, J. J., and Traylor, R. (**2003**). "A haptic back display for attentional and directional cueing," Haptics-e **3**, 1–20.

Terrence, P. I., Brill, J. C., and Gilson, R. D. (**2005**). "Body orientation and the perception of spatial auditory and tactile cues," *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, Orlando, FL.

van Erp, J. B. F., and Werkhoven, P. J. (**1999**). "Spatial characteristics of vibro-tactile perception on the torso," TNO-Rep. No. TM-99-B007, Soesterberg, The Netherlands, TNO Human Factors.

van Veen, H., and van Erp, J. (**2003**). "Providing directional information with tactile torso displays," in *Proceedings of EuroHaptics*, Dublin, Ireland.

Verrillo, R. T. (**1962**). "Investigation of some parameters of the cutaneous threshold for vibration," J. Acoust. Soc. Am. **34**, 1768–1773.

Verrillo, R. (**1963**). "Effect of contactor area on the vibrotactile threshold," J. Acoust. Soc. Am. **35**, 1962–1966.

Wall, C., Merfeld, D., Raucha, S., and Black, F. J. (**2002**). "Vestibular prostheses: The engineering and biomedical issues," J. Vestib. Res. **12**, 95–113.

Weisenberger, J. M. (**1994**). "Vibrotactile temporal masking: Effects of multiple maskers," J. Acoust. Soc. Am. **95**, 2213–2220.

Characteristics of whistles from resident bottlenose dolphins (*Tursiops truncatus*) in southern Brazil

Alexandre F. Azevedo[a)] and Alvaro M. Oliveira
Laboratório de Mamíferos Aquáticos (MAQUA), Depto. Oceanografia, Universidade do Estado do Rio de Janeiro, Rua São Francisco Xavier 524, Maracanã, Rio de Janeiro, 20550-013, Brazil

L. Dalla Rosa
Marine Mammal Research Unit, University of British Columbia, Room 215, Aquatic Ecosystems Research Laboratory, 2202 Main Mall, Vancouver, British Columbia, V6 T 1Z4, Canada

J. Lailson-Brito
Laboratório de Mamíferos Aquáticos (MAQUA), Depto. Oceanografia, Universidade do Estado do Rio de Janeiro, Rua São Francisco Xavier 524, Maracanã, Rio de Janeiro, 20550-013, Brazil

Little is known of the whistles produced by bottlenose dolphins in the South Atlantic Ocean. A total of 788 whistles were recorded from free-ranging bottlenose dolphins in Patos Lagoon estuary, southern Brazil. The mean number of whistles emitted per minute per animal was 0.8. Bottlenose dolphins emitted a varied repertoire of whistles, in which those with more than one inflection point were the most frequent and there was no predominance of ascending or descending whistles. Whistles recorded had a great frequency range, between 1.2 and 22.3 kHz. Whistle duration was 553.3 (±393.9 ms) and 66.6% of the whistles lasted <800 ms. Differences in the mean values of the whistles' characters were found between this study and other values previously reported for *Tursiops*. Bottlenose dolphins in the Patos Lagoon estuary emitted repeated whistle contours and individuals may be sharing some whistle types, as it has been suggested for *Tursiops*. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2713726]

I. INTRODUCTION

Bottlenose dolphins (*Tursiops* spp.) are cosmopolitan in distribution. *Tursiops truncatus* is found in most of the world's warm temperate to tropical seas, in coastal and offshore waters, while *T. aduncus* is limited to the coastal waters of the Indian and Western Pacific Ocean (Wells and Scott, 2002). The whistles of bottlenose dolphins, especially *T. truncatus*, have been studied extensively from captive or temporally restrained animals (e.g., Caldwell *et al.*, 1990; McCowan and Reiss, 2001; Watwood *et al.*, 2004) and free-ranging individuals (e.g., Sayigh *et al.*, 1990; Janik, 2000). These authors have reported mainly on signature whistle hypothesis, mimicry, and evidences for vocal learning.

Acoustic parameters of whistles, such as frequency and duration components and number of inflection points, have been used for characterization of whistles of delphinid species, allowing comparisons among groups and populations (Wang *et al.* 1995; Rendell *et al.* 1999; Bazúa-Duran and Au, 2004). However, only a few populations of free-ranging bottlenose dolphins have had their whistles characterized in terms of frequency and duration ranges, geographic variation, and others (e.g. Steiner, 1981; Schultz and Corkeron, 1994; Wang *et al.*, 1995; Morisaka *et al.*, 2005). Little is known of the whistles produced by bottlenose dolphins in the South Atlantic Ocean. The only published account by Wang *et al.* (1995), revealed values for ten variables of 110 whistles of bottlenose dolphins in Argentina. The authors reported frequency parameters ranging between 1.17 and 17.11 kHz and mean duration of 1.14 s (±0.49).

In southern Brazil, the Patos Lagoon estuary and adjacent coastal waters are inhabited by a small resident population of bottlenose dolphins, estimated at 83 individuals (95% CI=78 to 88) (Dalla Rosa, 1999). The dolphins are frequently found near the estuary mouth, and use the area for feeding, traveling, socializing, and resting (Möller, 1993). The average group size is four individuals (Dalla Rosa, 1999). Newborn calves are common in spring and summer (Möller, 1993). This paper describes the characteristics and presents sonograms of whistles recorded from free-ranging bottlenose dolphins in Patos Lagoon estuary.

II. METHODOLOGY

Acoustic recordings of underwater sound produced by bottlenose dolphins were made at Patos Lagoon estuary, southern Brazil (Fig. 1), between 4 and 7 March 2002. All surveys were carried out under similar weather conditions (Beaufort sea states ⩽2), in a small outboard-powered boat about 6 m in length. Dolphin group was an aggregation of two or more dolphins in apparent association within 100 m of each other. Trying to maximize data representativeness, we avoided oversampling groups/individuals during surveys, recording groups at different localities. Acoustic recordings were made with the engine off and were monitored by head-
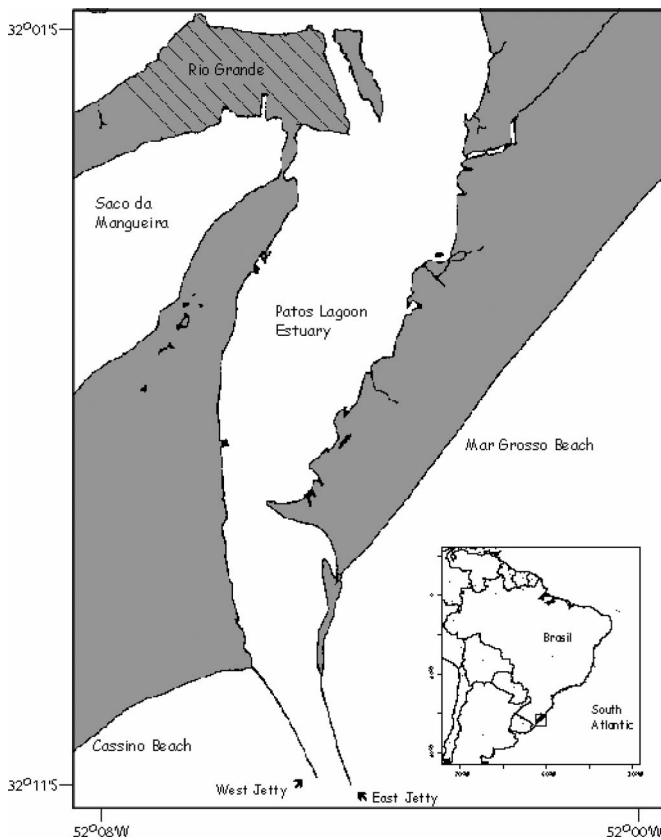
FIG. 1. Map of the Patos Lagoon Estuary, southern Brazil, where acoustic recordings of underwater sound produced by bottlenose dolphins were made.

phones. Whenever dolphin sounds became weak, we stopped recording and repositioned the boat. The recording system consisted of a High Tech Inc. hydrophone (model HTI-96-MIN, frequency response: 5 Hz to 30 kHz ±1.0 dB, −165 dB *re*: 1 V/μPa) and a digital audio tape recorder SONY TCD-D8 with upper frequency limit of 24 kHz (sampling rate of 48 kHz).

DAT recordings of each group sampled were redigitized using the program Cool Edit Pro (Syntrillium Software) at a sampling rate of 48 kHz, 16-bit resolution. Whistles were defined as continuous, narrow-band sound emissions with or without harmonics (Popper, 1980). Whistles were analyzed using Cool Edit Pro with a FFT size of 512 points, an overlap of 50%, and using a Hamming window.

The contour of each whistle was determined by visual analyses of the frequency modulation by at least two authors and was then categorized into the following broad classes: ascending (whistles rising in frequency and no one inflection point), descending (whistles falling in frequency and no one inflection point), ascending-descending (initial rising in frequency, one inflection point, then falling in frequency), descending-ascending (initial falling in frequency, one inflection point, then rising in frequency), constant (whistles in which the frequency changes 1000 Hz or less during more than 90% of duration), and multi (Fig. 2).

Seven acoustic parameters from the fundamental component of each whistle were measured: starting frequency (SF), ending frequency (EF), minimum frequency (MinF),

maximum frequency (MaxF), frequency range (MaxF-MinF), duration (DUR), and number of inflection points (defined as points where the whistle contour changed from ascending to descending or vice versa). The frequency variables were measured in kHz and the duration in milliseconds. We calculated the mean frequency (MeF) as the average of SF, EF, MinF, and MaxF. These whistle parameters were chosen to be consistent with previous studies of bottlenose dolphins (e.g., Wang *et al.*, 1995; Morisaka *et al.*, 2005) and other dolphin species (e.g., Bazúa-Duran and Au, 2004; Azevedo and Van Sluys 2005). We only used whistles for which all parameters of a spectral contour were distinctly measurable.

The descriptive statistics for all measured variables includes the minimum values, maximum values, means, standard deviation, and coefficient of variation. Over the whole set of whistles, distributions (Zar, 1999) were calculated for start frequency, end frequency, frequency range, and duration. The paired-sample *t* test (Zar, 1999) was applied to verify if the mean of the end frequency of all whistles analyzed was significantly different from the start frequency.

III. RESULTS

Eleven groups of bottlenose dolphins were recorded. Group size ranged from 2 to 15 members, including adults, juveniles, and mother-calf pairs. We estimate that about 40 different dolphins were recorded and, probably, some individuals were recorded more than one time. Animals were engaged in feeding, traveling, travel/feeding, and resting behaviors. A total of 982 whistles were recorded over 288 min, which represents 0.8 whistles per minute per individual. Of that total (N=982), 788 whistles had adequate signal quality for acoustical analysis. From the 788 whistles, 394 (50%) were tones with harmonics. Whistles with up to 14 inflection points were found, but those with zero up to four inflection points corresponded to 94% of all whistles.

Whistle duration was 553.3 ms (±393.9 ms) and 66.6% of the whistles lasted <800 ms. The average minimum frequency was 5.96±2.15 kHz and 83.6% of the whistles had MiF between 2.1 and 8.0 kHz. The maximum frequencies averaged 12.21±3.20 kHz and values between 12.1 and 16.0 kHz corresponded to 64.2%. The frequency range of whistles averaged 6.25±3.34 kHz. The average mean frequency was 8.70±2.15 kHz and 67.4% of the whistles had MeFs ranging from 6.1 to 10.0 kHz. Descriptive statistics of all measured whistle parameters are shown in Table I. The paired-sample *t* test (t=0.526; df=787; P=0.599) indicated that the end frequency (8.37±3.70) was not significantly different from the start frequency (8.28±3.11).

Whistles categorized as multi (more than one inflection point) were the most common and corresponded to 31.5% of all whistles. Ascending-descending (23.5%), ascending (17.3%), and descending (14.2%) whistles also were frequently emitted. Descending-ascending (7.4%) and constant (6.1%) whistles were less frequent. Descriptive statistics for acoustic parameters of bottlenose dolphin's whistles for each whistle category are shown in Table II.

A)Ascending
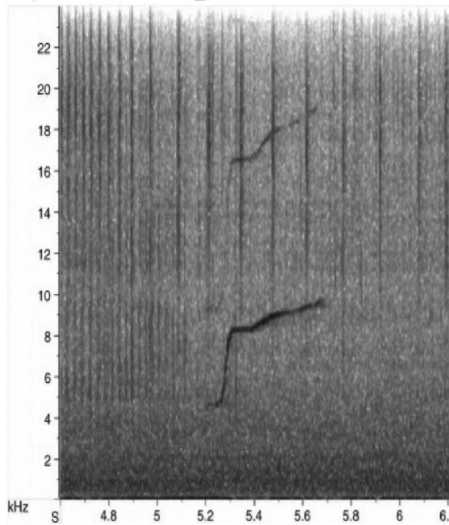
B)Descending
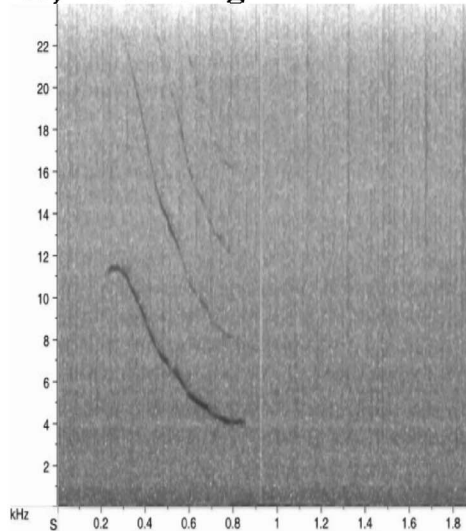
C)Ascending-Descending
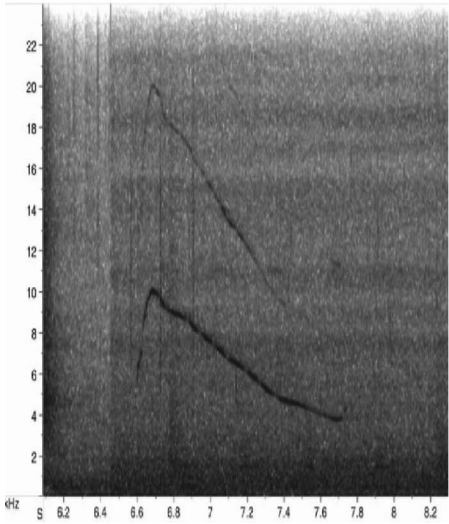
D)Descending-Ascending
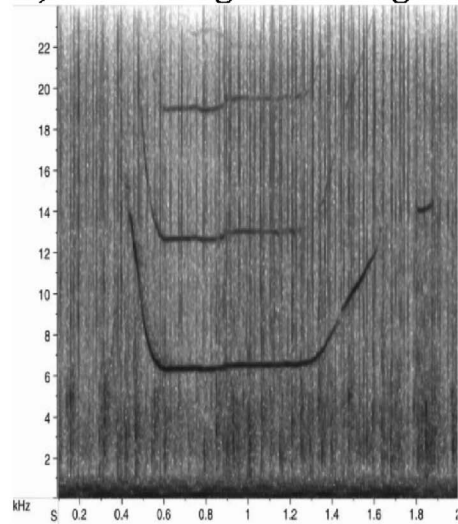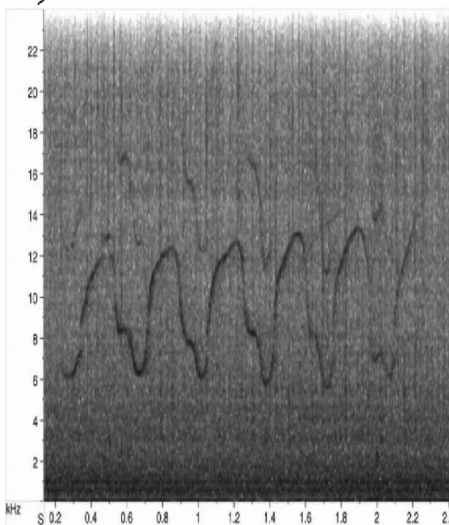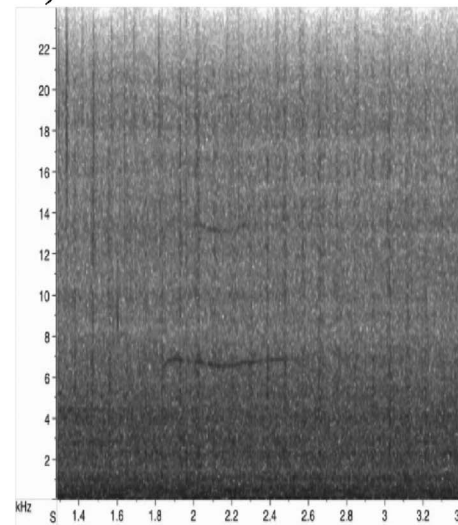
FIG. 2. The six classes to which bottlenose dolphin whistles contours were categorized. X axis=time (s), Y axis=frequency (kHz).

E)Multi

F)Constant

TABLE I. Descriptive statistics for acoustic parameters of bottlenose dolphin whistles in the Patos Lagoon estuary, southern Brazil (*N*=788). The frequency variables were measured in kHz and the duration in milliseconds.

Acoustic parameters	Range	Mean (±S.D.)	Coefficient of variation (%)
Starting frequency	3.1–20.8	8.28 (3.11)	37.6
Ending frequency	2.8–22.3	8.37 (3.70)	44.2
Minimum frequency	1.2–17.2	5.96 (2.15)	36.1
Maximum frequency	3.6–22.3	12.21 (3.20)	26.2
Frequency range	0.1–16.6	6.25 (3.34)	53.4
Mean frequency	3.5–18.0	8.70 (2.15)	24.7
Duration	48–2458	553.3 (393.9)	71.2
Inflections	0–14	1.42 (1.85)	92.5

Similar whistle contours were found in 101 whistles, totaling 435 whistles (55.2%) with at least one repetition. Whistle contours were repeated in different groups and different days. The most repeated whistle contours occurred 25 times, and seven had at least ten repetitions.

IV. DISCUSSION AND CONCLUSIONS

Bottlenose dolphins in Patos Lagoon estuary emitted a varied repertoire of whistles, in which those with more than one inflection point were the most frequent. This is in accordance with previous studies of free-ranging bottlenose dolphins, which reported mean inflection points between 0.78 (Morisaka *et al.*, 2005) and 2.86 (Steiner, 1981). Whistles recorded in our study had a great frequency range, but, in general, were similar to published frequency ranges for this genus. Schultz and Corkeron (1994) reported bottlenose dolphin whistles with a low frequency of 0.8 kHz in Morenton Bay (Australia), and Wang *et al.* (1995) found high frequency reach up to 21.6 kHz. The range of duration of bottlenose dolphin whistles sampled was also in agreement with previous studies. Duration generally varies greatly among bottlenose whistles, with changes as great as 0.37 s (Morisaka *et al.*, 2005) to 1.30 s (Wang *et al.*, 1995). Both

parameters, duration, and inflection points had the highest coefficients of variation. This high intraspecific variability in both duration and number of inflection points may result from an individual modulation of these parameters so that information may be transmitted from different contexts or from different individuals (Steiner, 1981; Wang *et al.*, 1995; Rendell *et al.*, 1999; Bazúa Duran and Au, 2004).

Some whistle characters observed in this study differed significantly from those values previously reported for *Tursiops* spp. (Table III). The duration of whistles recorded at Patos Lagoon estuary differed from all areas, except for one location in Japan. Comparisons of the number of inflection points and frequency variables showed similarities and differences with other populations of *T. truncatus* and *T. aduncus*. Some studies have indicated that species-specific variables have low intraspecific and high interspecific variation (Steiner, 1981; Wang *et al.*, 1995), such that we would expect to observe similarities with *T. truncatus* rather than with *T. aduncus*. These species are closely related (Rice, 1998), potentially explaining the increase in intraspecific variability for some comparisons. Other, nonheritable, factors also may be causing increased intraspecific variability. Variation of acoustic whistle parameters may be related to adaptation to background noise (Rendell *et al.*, 1999). For example, Wang *et al.* (1995) suggested that in bottlenose dolphin whistles, higher frequencies, longer durations, and greater numbers of inflections are associated with localities of higher background noise. Additionally, social relationship and behavioral states at recording time may also be responsible for differences and similarities among the studies. Spinner dolphins show high variation in whistle duration (Bazúa Duran and Au, 2002), which might be attributed to differences in group size and general behavioral states.

Bottlenose dolphins in the Patos Lagoon estuary emitted repeated whistle contours. Several authors (e.g., Caldwell *et al.*, 1990; Sayigh *et al.*, 1990; Janik, 2000; Watwood *et al.*, 2004) have suggested that the production of repeated whistles by bottlenose dolphins may indicate the use of sig-

TABLE II. Minimum, maximum, mean, and standard deviation values for acoustic parameters of bottlenose dolphin whistle categories in the Patos Lagoon estuary, southern Brazil. The frequency variables were measured in kHz and the duration in seconds.

Parameters	Ascending (*N*=137)	Descending (*N*=112)	Ascending-descending (*N*=185)	Descending-ascending (*N*=58)	Constant (*N*=48)	Multi (*N*=248)
Duration	0.05–1.15	0.07–1.07	0.12–1.29	0.2–1.53	0.05–0.82	0.18–2.46
	0.30±0.19	0.41±0.20	0.49±0.22	0.44±0.28	0.26±0.18	0.88±0.47
Starting frequency	3.2–12.9	4.4–20.8	3.6–16.5	3.9–17.1	3.5–14.4	3.1–17.7
	6.9±2.30	11.07±3.15	8.1±2.72	9.4±3.20	6.9±3.01	7.9±2.92
Ending frequency	6.0–20.4	3.2–14.2	2.8–15.6	3.8–22.3	3.5–14.5	2.9–21.0
	12.41±2.86	5.34±1.64	7.2±3.26	9.5±3.70	7.0±2.99	8.5±3.37
Minimum frequency	3.2–12.9	3.2–14.2	1.4–12.9	3.2–12.3	3.4–14.0	1.2–17.2
	6.9±2.30	5.34±1.64	6.1±2.09	6.0±2.16	6.6±3.04	5.6±1.90
Maximum frequency	6.0–20.4	4.4–20.8	5.8–20.3	6.8–22.3	3.6–14.5	7.1–21.0
	12.41±2.86	11.07±3.15	13.2±2.74	11.4±3.11	7.2±3.03	13.1±2.66
Mean frequency	4.7–15.0	4.1–15.2	4.5–13.4	5.5–15.7	3.5–14.2	4.7–17.9
	9.5±2.19	8.2±1.88	12.6±1.94	9.1±2.33	6.9±3.01	8.8±1.88
Frequency range	1.1–12.8	2.8–16.6	1.2–15.2	1.5–15.3	0–1.0	1.1–15.4
	5.3±2.80	5.9±3.17	7.1±3.03	5.4±2.99	0.6±0.32	7.6±2.93
Inflections	0	0	1	1	0	2–14
						3.5±2.06

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Azevedo *et al.*: Bottlenose dolphins whistles in southern Brazil   2981

TABLE III. Mean and standard deviation of some whistle parameters of bottlenose dolphins from previous studies. The two-sided *t*-test (Zar. 1999) was performed to compare whistle parameters with this study. Italic numbers represent significant differences ($P < 0.01$). An asterisk indicates data not reported by the authors.

Location	SF	EF	MinF	MaxF	DUR	I	N	Study
Patos Lagoon estuary, Brazil[a]	8.28 (3.11)	8.37 (3.70)	5.96 (2.15)	12.21 (3.20)	0.55 (0.39)	1.42 (1.85)	788	This study
Argentina[a]	*9.24 (2.74)*	*6.63 (2.29)*	*5.91 (1.50)*	*13.65 (1.54)*	*1.14 (0.49)*	1.58 (1.24)	110	Wang *et al.* (1995)
Texas, USA[a]	8.01 (2.81)	8.16 (3.78)	5.77 (1.84)	*11.32 (3.31)*	0.68 (0.40)	2.09 (2.54)	2022	Wang *et al.* (1995)
North Atlantic Ocean[a]	*11.26 (3.99)*	*10.20 (3.65)*	*7.33 (1.66)*	*16.24 (2.69)*	*1.30 (0.63)*	2.86 (2.45)	858	Steiner (1981)
Sado estuary, Portugal[a]	*5.8 (1.8)*	*12.1 (4.4)*	*15.0 (2.7)*	*5.4 (1.2)*	*0.86 (0.40)*	*	735	dos Santos *et al.* (2005)
Gulf of California[a]	*12.10 (2.89)*	*9.19 (3.44)*	*6.91 (2.11)*	*13.68 (1.72)*	0.66 (0.35)	1.15 (1.32)	110	Wang *et al.* (1995)
Eastern Tropical Pacific Ocean[a]	*11.2 (4.6)*	*9.0 (3.7)*	*7.4 (2.2)*	*17.2 (3.1)*	*1.4 (0.7)*	*3.7 (3.0)*	157	Oswald *et al.* (2003)
Moreton Bay, Australia[b]	*	*	*	*	0.38 (0.21)	*	404	Schultz and Corkeron (1994)
Shark Bay, Australia[b]	*3.84 (1.42)*	*7.56 (3.80)*	*3.57 (0.97)*	10.57 (3.02)	0.68 (0.35)	1.63 (1.53)	658	Wang *et al.* (1995)
Japan[b]	*10.33 (2.41)*	8.87 (2.21)	*7.37 (1.54)*	11.62 (2.00)	0.62 (0.34)	*0.88 (0.79)*	215	Wang *et al.* (1995)
Mikura I., Japan[b]	*7.17 (2.85)*	*9.82 (4.18)*	5.98 (2.44)	12.21 (3.20)	*0.39 (0.33)*	1.22 (1.39)	851	Morisaka *et al.* (2005)
Ogasawara I., Japan[b]	*6.91 (3.12)*	*10.35 (4.86)*	5.61 (2.06)	12.34 (4.93)	*0.44 (0.44)*	1.19 (1.50)	247	Morisaka *et al.* (2005)
Amakura-Shimoshima I., Japan[b]	*6.74 (2.82)*	8.06 (3.80)	5.63 (2.21)	*9.39 (3.90)*	*0.37 (0.25)*	*0.78 (0.88)*	515	Morisaka *et al.* (2005)

[a]Tursiops truncatus.
[b]Tursiops aduncus.

nature whistles, although this hypothesis is not wholly supported (McCowan and Reiss, 2001). Like in other populations, bottlenose dolphins from the Patos Lagoon estuary live in fission-fusion societies where individual association patterns are fluid (Dalla Rosa, 1999). Therefore, signature whistles could be used to individual recognition or group cohesion (e.g., Caldwell *et al.*, 1990; Sayigh *et al.*, 1990; Janik and Slater, 1998; Watwood *et al.*, 2004). However, our method of data collection did not allow us to identify the whistler, therefore we cannot evaluate the signature whistle hypothesis with this data set.

It has been suggested that bottlenose dolphins imitate the whistles of conspecifics (Janik, 2000). But whistles produced by bottlenose dolphins may come from a common shared whistle repertoire and what appears to be imitation simply refers to animals repeating the same call type, what might serve as contact calls or to coordinate group movements or group formation (Watwood *et al.*, 2004). In this way, individuals in the Patos Lagoon estuary may be sharing some whistle types, as it has been suggested for *Tursiops* elsewhere.

This is the first description of the whistle repertoire of bottlenose dolphins in Brazilian waters and the second in the South Atlantic Ocean. We recorded and analyzed whistles from free-ranging bottlenose dolphins engaged in different behaviors. The whistle repertoire of bottlenose dolphins in the Patos Lagoon estuary seems to be varied, as it has been observed for *Tursiops* in other areas (e.g., dos Santos *et al.*, 2005). Bottlenose dolphins from the Patos Lagoon estuary produced stereotyped whistles in which the acoustic parameters were similar to published ranges for the genus. But, some differences were found between whistle characters of bottlenose dolphins from Patos Lagoon estuary and those values previously reported for *Tursiops* spp. Besides biological differences, the comparisons with previously published whistle characteristics may have been affected, in part, by methods and equipment used to record dolphin sounds in

each study. So, care must be taken in order to avoid speculative conclusion about intra- and interspecific variability in *Tursiops* whistles.

There was no predominance of ascending or descending whistles and the mean of the SF and EF values were similar. The relationships between start and end frequency may vary between bottlenose populations (Wang *et al.*, 1995; dos Santos *et al.*, 2005). The statement that bottlenose whistles in Patos Lagoon estuary are mainly balanced in frequency represents an important characteristic of this population and, consequently, may be a tool to discriminate *Tursiops* populations along the Brazilian coast. Additionally, production of repeated whistles was verified in these free-ranging dolphins, but further studies are needed to clarify its function.

ACKNOWLEDGMENTS

Azevedo, A. F., and Van Sluys, M. (**2005**). "Whistles of tucuxi dolphins (*Sotalia fluviatilis*) in Brazil: comparisons among populations," J. Acoust. Soc. Am. **117**, 1456–1464.
Bazúa-Duran, M. C., and Au, W. W. L. (**2002**). "Whistles of Hawaiian spinner dolphins," J. Acoust. Soc. Am. **112**, 3064–3072.
Bazúa-Duran, M. C., and Au, W. W. L. (**2004**). "Geographic variations in the whistles of spinner dolphins (*Stenella longirostris*) of the Main Hawaiian Islands," J. Acoust. Soc. Am. **116**, 3757–3769.
Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). "Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego), pp. 199–234.

Azevedo *et al.*: Bottlenose dolphins whistles in southern Brazil

Dalla Rosa, L. (**1999**). "Estimativa do tamanho da população de botos, *Tursiops truncatus*, do estuário da Lagoa dos Patos, RS, a partir da foto-identificação de indivíduos com marcas naturais e da aplicação de modelos de marcação-recaptura" ("Population size of bottlenose dolphin, *Tursiops truncatus*, in the Patos Lagoon estuary, RS, using photo-id and mark-recapture models"), Master thesis, Fundação Universidade do Rio Grande, Rio Grande, Brazil.

dos Santos, M. E., Louro, S., Couchinho, M., and Brito, C. (**2005**). "Whistles of Bottlenose Dolphins (Tursiops truncatus) in the Sado Estuary, Portugal: Characteristics, Production Rates, and Long-Term Contour Stability," Aquat. Mamm. **31**(4), 453–462.

Janik, V. M. (**2000**). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," Science **289**, 1355–1357.

Janik, V. M., and Slater, P. J. B. (**1998**). "Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls," Anim. Behav. **56**, 829–838.

McCowan, B., and Reiss, D. (**2001**). "The fallacy of 'signature whistles' in bottlenose dolphins: a comparative perspective of 'signature information' in animal vocalizations," Anim. Behav. **62**, 1151–1162.

Möller, L. M. (**1993**). "Observações sobre o comportamento e a ecologia do boto *Tursiops truncatus* no estuário da Lagoa dos Patos, RS, Brasil" ("Observations about behavior and ecology of the bottlenose dolphin *Tursiops truncatus* in the Patos Lagoon estuary, RS, Brazil"), Master thesis, Fundação Universidade do Rio Grande, Rio Grande, Brazil.

Morisaka, T., Shinohara, M., Nakahara, F., and Akamatsu, T. (**2005**). "Geographic variations in the whistles among three Indo-pacific bottlenose dolphin *Tursiops aduncus* populations in Japan," Fish. Sci. **71**, 568–576.

Oswald, J. N., Barlow, J., and Norris, T. F. (**2003**). "Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean," Marine Mammal Sci. **19**, 20–37.

Popper, A. N. (**1980**). "Sound emission and detection by delphinids," in *Cetacean Behaviour: Mechanisms and Functions*, edited by L. M. Herman (Wiley, New York), pp. 1–52.

Rendell, L. E., Matthews, J. N., Gill, A., Gordon, J. C. D., and MacDonald, D. W. (**1999**). "Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation," J. Zool. Lond. **249**, 403–410.

Rice, D. W. (**1998**). *Marine Mammals of the World: Systematics and Distribution* (The Society for Marine Mammalogy, Lawrence).

Sayigh, L. S., Tyack, P. L., Wells, R. S., and Scott, M. D. (**1990**). "Signature whistles of free-ranging bottlenose dolphins *Tursiops truncatus*: Stability and mother-offspring comparisons," Behav. Ecol. Sociobiol. **26**, 247–260.

Schultz, K. W., and Corkeron, P. J. (**1994**). "Interspecific differences in whistles produced by inshore dolphins in Moreton Bay, Queensland, Australia," Can. J. Zool. **72**, 1061–1068.

Steiner, W. W. (**1981**). "Species-specific differences in pure tonal whistle vocalizations of five western North Atlantic dolphin species," Behav. Ecol. Sociobiol. **9**, 241–246.

Wang, D., Wursig, B., and Evans, W. E. (**1995**). "Whistles of bottlenose dolphins: Comparisons among populations," Aquat. Mamm. **21**, 65–77.

Watwood, S. L., Tyack, P. L., and Wells, R. S. (**2004**). "Whistle sharing in paired male bottlenose dolphins, *Tursiops truncatus*," Behav. Ecol. Sociobiol. **55**, 531–543.

Wells, R. S., and Scott, M. D. (**2002**). "Bottlenose Dolphins," in *Encyclopedia of Marine Mammals*, edited by W. F. Perrin, B. Wursig, and J. G. M. Thewissen (Academic, San Diego), pp. 122–128.

Zar, J. H. (**1999**). *Biostatistical Analysis* (Prentice Hall, Englewood Cliffs, NJ).

Acoustic signals of Chinese alligators (*Alligator sinensis*): Social communication

Xianyan Wang
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan 430072, People's Republic of China and Graduate School of the Chinese Academy of Sciences, Beijing 100039, People's Republic of China

Ding Wang[a)]
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan 430072, People's Republic of China

Xiaobing Wu
College of Life Sciences, Anhui Normal University, Wuhu 241000, People's Republic of China

Renping Wang and Chaolin Wang
Anhui Research Center for Chinese Alligator Reproduction, Xuancheng 242034, People's Republic of China

This paper reports the first systematic study of acoustic signals during social interactions of the Chinese alligator (*Alligator sinensis*). Sound pressure level (SPL) measurements revealed that Chinese alligators have an elaborate acoustic communication system with both long-distance signal—bellowing—and short-distance signals that include tooting, bubble blowing, hissing, mooing, head slapping and whining. Bellows have high SPL and appear to play an important role in the alligator's long range intercommunion. Sounds characterized by low SPL are short-distance signals used when alligators are in close spatial proximity to one another. The signal spectrographic analysis showed that the acoustic signals of Chinese alligators have a very low dominant frequency, less than 500 Hz. These frequencies are consistent with adaptation to a habitat with high density vegetation. Low dominant frequency sound attenuates less and could therefore cover a larger spatial range by diffraction in a densely vegetated environment relative to a higher dominant frequency sound. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2714910]

I. INTRODUCTION

Acoustic signals are widely used by insects (Banks and Thompson, 1985; Robert *et al.*, 1995; Snedden, 1996; Janne *et al.*, 2004), anurans (Ryan, 1991; Gerhardt, 2001; Sergio *et al.*, 2003; Tobias *et al.*, 2004), birds (Gibson and Bradbury, 1985; Douglass, 1989; Lind *et al.*, 1996; Liu *et al.*, 1998; Valentin *et al.*, 2004), primates (Marler, 1973; Brown, 1989; Snowdon and Elowson, 1999; Stella and Charles, 2002) and many other mammals (Brandon *et al.*, 2003; Susan and Gerald, 2004; Volker *et al.*, 2005). Reptiles are relatively nonvocal. Acoustic signaling in reptiles has consequently received little attention (Marcellini, 1977; Vliet, 1989; Young, 1991; Roberto *et al.*, 2003). In contrast, crocodilians are quite vocal relative to other reptile groups and the alligators are among the most vocal of the crocodilians (Campbell, 1973; Gans and Maderson, 1973; Garrick, 1975; Garrick and Lang, 1977; Herzog and Burghardt, 1977; Chen *et al.*, 1985). Garrick *et al.* (1978) suggested that the highly vocal nature of alligators might be an adaptation to communication in the

marsh habitats in which they live, which are dense and thus restrict the transmission of visual signals in social communication.

The Chinese alligator, *Alligator sinensis*, is a small crocodilian species within the family *Alligatoridae*. It is one of the most endangered of the 23 crocodilian species worldwide (Thorbjarnarson, 1992). Wild individuals currently number less than 150 and are restricted to a small area in southern Anhui Province and perhaps in adjacent Zhejiang and Jiangsu Provinces (Ding *et al.*, 2001; Chen *et al.*, 2003). The communication systems of Chinese alligators are quite complex, involving stimuli to several sensory apparatus, including auditory, visual and olfactory cues (Chen *et al.*, 1985, 2003). Vocal behavior is an important component of the behavioral repertoire of Chinese alligators (Zhang *et al.*, 2005). Unlike the well-studied American alligator, *A. mississippiensis*, which is the only other member of the genus *Alligator*, acoustic communication of *A. sinensis* has only been described based on direct observational data. Zhu (1957) mentioned the thunder-like roaring vocalization produced by the Chinese alligator, and Garrick (1975) reported the structure and pattern of the roars. Chen *et al.* (1985, 2003) found that both male and female Chinese alligators bellow to attract mates during the courtship season. They also noted that alligators produced different vocalizations when frightened, dis-

---
[a)]Author to whom correspondence should be addressed. Electronic mail: wangd@ihb.ac.cn

tressed and aggressive. Wu and Wang (2004) and Wang *et al.* (2006) monitored the timing of bellowing in wild and semi-natural conditions and found that Chinese alligators bellow throughout their active season, with the highest frequency during the courtship season.

However, little detailed analysis of the acoustic signals of Chinese alligators has been done. This study presents the first detailed analysis of the acoustic signals used in Chinese alligators social communication. Field recordings of the vocalizations were conducted at the Anhui Research Center for Chinese Alligator Reproduction (ARCCAR), together with related vocal behaviors. Temporal and frequency characteristics of Chinese alligator acoustic signals were described using spectrographic analyses.

II. MATERIALS AND METHODS

A. Study site and subjects

Data were collected during May–October 2005 and May–June 2006 in the reproductive area (30°54'N, 118°46'E) at ARCCAR, located in Xuancheng, Anhui, China. The study site is a 1.6 ha. enclosure that primarily consists of four permanent ponds. We chose pond No. 3 as the focal study locality. Its surface area is approximately 2000 m² with a 400 m² island situated in the center covered in dense shrubbery [for a detailed description of the study area see Wang *et al.* (2006)]. Vegetation in the study area is mostly horse-tailed pine (*Pinus massoniana*), bitter bamboo (*Pleioblastus amarus*), weeping willow (*Salix babylonica*) and common reed (*Phragmites communis*) (Chen *et al.*, 1985; Zhou *et al.*, 2004). There are 120 Chinese alligators living in the enclosure. Most individuals are wild-caught, with a few that are the first and second generations produced from artificial reproduction. The male: female ratio of the reproductive cohort was 1:4 [see Zhang *et al.* (2005) for more details].

B. General observations and SPL measurements

Terms that had been used to describe the bellowing behavior of the American alligator were employed to describe the vocal behaviors of the Chinese alligator (e.g., bout and chorus, see below). We collected information on vocalizations, calls in each bout and number of bouts, using the "target objective observation" method (Zhang *et al.*, 2005). Observations were taken from behind a brick boundary wall (1.5 m high) by naked eyes or Fujinon 7×50 binoculars to avoid disturbance to the animals. Observational distances ranged from 2 to 30 m.

The sound pressure level (SPL, in dB) of all acoustic signals was measured by pointing a hand-held Realistic sound level meter (model DT-805, re. 20 μPa, C weighting, fast response) with a windshield cover towards the vocalizing animals. The measurements were conducted at a distance of 5 m except for "tooting" and "hissing" which were performed at 2 m because these sounds were too weak to be measured at greater distances (i.e., the SPL is equal to background noise >2 m distance).

C. Sound recording and analysis

We designed a portable signal recording system, consisting of an omnidirectional microphone (Fidek KSM-81, frequency response: 30 Hz–20 kHz, +5.5/−11.4 dB, Japan) connected to a Sony digital tape recorder (TCD-D100, sampling frequency: 48 kHz, Japan) and Sony digital audio tape. The system allowed us to record signals at a controlled distance from vocalizing animals. When required for quality recordings we fixed the microphone to one end of a 5 m bamboo pole and extended the pole towards vocalizing animals. Sound recordings were made during the chorus (bellow) or opportunistically (other signals) on calm days without wind to avoid microphone saturation. The recording distance of each signal was the same as for the SPL measurements.

Recording analyses were performed using SIGNAL/RTS™ (version 4.0, 2006, American Engineering Design) software, a PC-based signal processing system. Recordings were replayed from the tape recorder and digitized by a 12 bit Data Translation-2821G analog-to-digital board with a sampling frequency of 25 kHz. We manually selected signals that had high signal-to-noise ratio. We then calculated the four acoustic variables (dominant frequency, call duration, note and harmonic, see below) of each signal using wave form and spectrogram analysis (fast Fourier transform, the 256 points option, and Hanning's sampling window).

To interpret the bioacoustics of vocal behaviors and acoustic signals, we used the following definitions [adapted from Duellmann (1970), Garrick *et al.* (1978), Duellmann and Pyles (1983), Vliet (1989) and Li *et al.* (2005)]:

Bout: a succession of vocalizaions/calls produced in a regularly repeated unit.

Chorus: the combined vocalizing bouts of all individuals produced in a certain time frame.

Call duration: time between two points at which the sound wave oscillations rose from the background noise and descended into the background noise.

Note: temporally uninterrupted sound element composing the call, consisting of a group of pulses.

Dominant frequency: frequency at which a call or note presents the highest intensity (energy), measured in Hz or kHz.

Fundamental frequency: the lowest frequency component in voiced speech sounds.

Harmonic: an integral multiple of the fundamental frequency.

All of the acoustic characters were fed into SPSS 13.0 software for subsequent analysis. All values are reported as mean ± SE (*n*=sample size).

III. RESULTS

From the field observations and SPL measurements, signals were divided into two categories (Fig. 1): (A) long-distance signal—bellowing, characterized by high SPL and covering a large spatial range; and, (B) short-distance signals, including tooting, bubble blowing, hissing, mooing,

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Wang *et al.*: Acoustic signals of Chinese alligators   2985
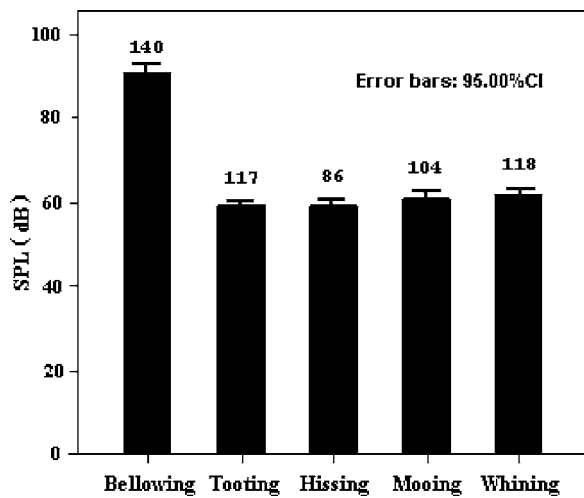
FIG. 1. SPL (in dB) of acoustic signals in Chinese alligator social communication. Long-distance signal (bellowing) had a higher SPL than short-distance signals (tooting, hissing, mooing and whining). Head slapping and bubble blowing were omitted here because they are produced by slapping or blowing water. (Mann-Whitney U-Test, $U=-17.76$, $P<0.001$, two-tailed). Numbers above bars indicate sample sizes.

head slapping and whining, which have low SPL and mostly occurred when alligators were in close spatial proximity to each other.

A. Long-distance signal
Bellowing

Alligators bellowed almost daily during the courtship season. One alligator bellowing stimulated others to crawl into the bellowing pond and bellow synchronously, eventually forming a bellowing chorus. Alligators bellowed with a "head oblique tail arched" posture, their heads held oblique at a 50–80° angle to the water surface and their bodies submerged except for their tails that were arched in a "bow" formation above the water (Fig. 2). During the chorus vocalizations, the alligators remained immobile and bellowed in one direction.

FIG. 2. Posture of the bellowing activity of the Chinese alligator, with a head oblique tail arched posture.
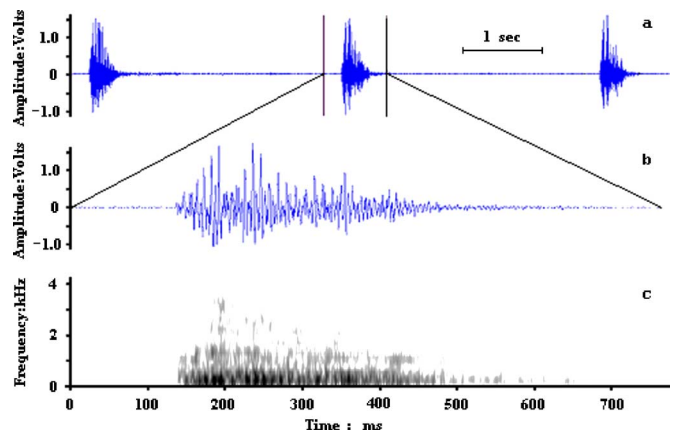
FIG. 3. (Color online) Wave form and spectrogram of the long-distance signal—bellowing. (a) Wave form of a bout with three bellows; (b) and (c) show the wave form and spectrogram of the second bellow of the bout, respectively.

Choruses lasted 10.4 ± 0.25 min ($n=79$), with 5.86 ± 0.32 bouts ($n=95$), 3.32 ± 0.79 bellows ($n=130$) in each bout, for a total of 14.3 ± 0.66 bellows per alligator ($n=92$). The temporal and frequency features of bellows are shown in Table I and Fig. 3. The bellows had an average SPL of 90.8 ± 0.34 dB ($n=140$; Fig. 1) at 5 m.

B. Short-distance signals
1. Tooting and bubble blowing

When two or more alligators met at the water surface, they would toot or blow bubbles for a moment. Toots were produced from the nares with the mouth underwater, and bubbles were produced by blowing water. These behaviors typically occurred prior to head rubbing, which is a stereotypical behavior of Chinese alligators before copulation. The temporal and frequency features of tooting are summarized in Table I and Fig. 4(a). The toots had an average SPL of 59.1 ± 0.31 dB ($n=117$) (Fig. 1) at 2 m.

2. Hissing

Hissing mostly occurred when aggressors, humans or other alligators, intruded by approaching too closely, especially in proximity to nest boxes protected by mother alligators. Hisses were produced from the nares with the mouth closed. A hiss is divided into two distinct notes [Fig. 4(b)]. The first is associated with an expiration about 4874 ± 64.1 ms in duration and, after a 255 ± 5.56 ms pause, the hiss continues with an inspiration about 5175 ± 70.7 ms in duration ($n=98$). Thus the entire hiss lasted $10,306\pm119$ ms. The pattern was repeated several times until the aggressor departed. Table I and Fig. 4(b) show the temporal and frequency features of the hiss. Hisses had an average SPL of 59.0 ± 0.36 dB ($n=86$; Fig. 1) at 2 m.

3. Mooing and head slapping

Moos were produced with a low "head oblique tail arched" posture. An average of 7.50 ± 1.75 moos ($n=26$) occurred in each mooing bout. Moos were followed by an average of 6.38 ± 2.06 head slappings ($n=26$). The most com-

TABLE I. Temporal and frequency features of acoustic signals in Chinese alligator social communication.

	Dominant frequency (Hz)	Call duration (ms)	Notes per call	Harmonics
Bellowing	201±3.77 *n*=126	560±7.06 *n*=133	1	⋯
Tooting	186±4.99 *n*=147	217±4.89 *n*=154	3–5	⋯
Hissing	194±4.15 *n*=93	10 306±119 *n*=98	2	⩾3
Mooing	273±5.59 *n*=104	1029±24.3 *n*=120	2	⋯
Whining	236±5.28 *n*=101	361±8.08 *n*=107	1	⋯

mon response of alligators receiving mooing and head slapping was swimming away or submerging. The temporal and frequency features of the moo are shown in Table I and Fig. 4(c). Moos had an average SPL of 60.6±0.40 dB (*n* =104; Fig. 1) at 5 m.

4. Whining

Whines were produced by relatively small alligators. When chased or attacked by larger alligators, the smaller ones would run away or submerge. The temporal and frequency features of whining are shown in Table I and Fig. 4(d). Whines had an average SPL of 61.4±0.35 dB (*n*=118; Fig. 1) at 5 m.

IV. DISCUSSION

Chinese alligators have a complex acoustic communication system. They exhibit both long-distance and short-distance acoustic signals, a phenomenon previously reported in songbirds (Liu *et al.*, 1998; Selvino and Carel, 2004) and nonhuman primates (Marler, 1973; Stella and Charles, 2002).

Bellowing, the long-distance signal, is characterized by a high SPL and covers a large spatial range. Because Chinese alligators are solitary (Chen *et al.*, 1985, 2003) and must aggregate to copulate during the mating season (Wang *et al.*, 2006), a signal such as bellowing that can be heard over a large range is required for alligators to assemble. Wang *et al.* (2006) found that bellowing played an important role in assembling individuals into ponds to copulate during the breeding season.

Compared to visual and olfactory signals, auditory signals are an effective way for alligators to aggregate. Dense vegetation restricts visual communication. A chemical pheromone, emitted from pairs of musk glands medial to the mandible on the underside of the jaw and the lateral walls of the cloaca (Chen *et al.*, 1985, 2003), is easily dispersed by wind but the direction of spread cannot be controlled by the signaler. This lack of control reduces the efficacy of olfactory
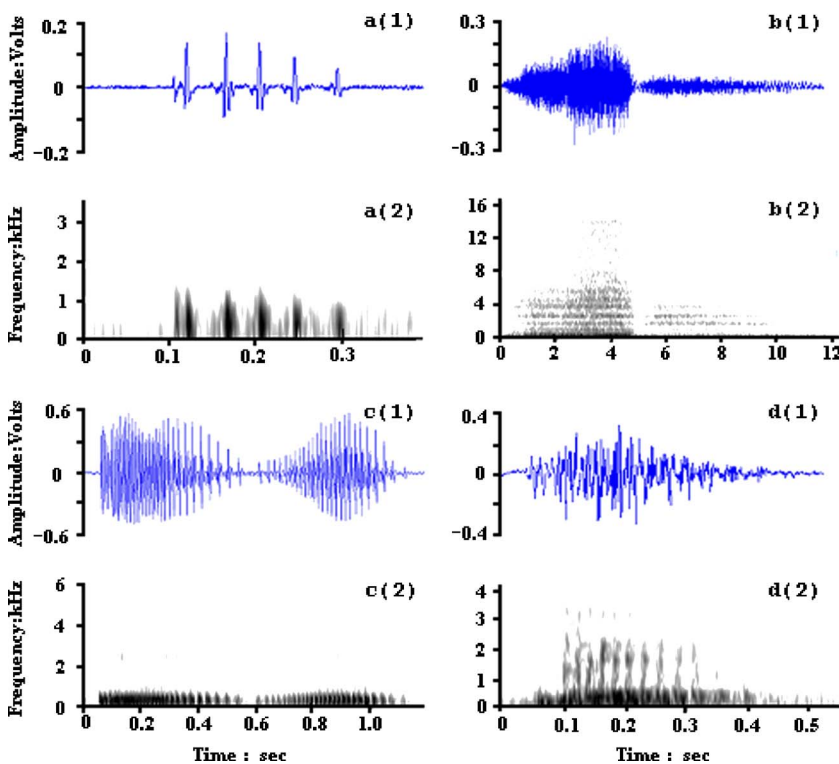
FIG. 4. (Color online) Wave form (1) and spectrogram (2) of the short-distance signals. Panels (a), (b), (c) and (d) show wave forms and spectrograms of tooting, hissing, mooing, and whining, respectively (head slapping and bubble blowing were omitted here because they are produced by slapping or blowing water).

signaling to aggregate the lek. Given the short courtship season (Chen *et al.*, 1985, 2003) and the low aspect of the alligator body, auditory cues are likely to be energetically more efficient relative to visual or olfactory cues for attracting potential mates. Bellows also indicate the availability of water necessary for copulations.

Short-distance signals (low SPL) include tooting, bubble blowing, hissing, mooing, head slapping and whining. These typically occurred when alligators were in close proximity to each other. Short-distance signals are diverse, are used in different behavioral contexts and have varied temporal properties (e.g., call duration, numbers of notes per call, etc.). The different signals may convey different information, though what this information is and its evolutionary function requires more investigation.

Vocalizing is energetically costly. Optimizing the propagation efficiency of acoustic signals is therefore crucial for saving energy. The environment can affect the amplitude and frequency of a call such that it can degrade signal effectiveness and make it unrecognizable to the receiver (review in Bosch and Riva, 2004). The acoustic adaptation hypothesis predicts that acoustic signals will be shaped via natural selection to display maximum propagation within the habitat in which the signals evolved (Morton, 1975). The environmental selection hypothesis predicts that physical characteristics of the environment influence the evolution of acoustic signals by favoring those properties that decrease sound attenuation and distortion with distance (Sergio *et al.*, 2003; Bosch and Riva, 2004). Thus a species' acoustic signals should be matched to its habitat with regards to efficiency of signal transmission.

The very low dominant frequencies shown by our spectrographic analysis (Table I) are consistent with adaptation to high density vegetation. Lower frequency sounds attenuate less relative to higher frequency sounds, especially in dense vegetation, and can travel longer distances. This is because they are less easily absorbed than higher frequency sounds (Bosch and Riva, 2004). Alligators produce sounds near the ground or water surface, which can cause ground attenuation due to interference between direct and ground-reflected waves (Marten *et al.*, 1977). Potential attenuation can be mitigated by producing sounds using different postures. Bellows are produced with the heads oblique at a 50–80° angle to the surface of water (Fig. 2), which may facilitate sound propagation to a higher height. Because short-distance signals have low SPL and occur when alligators are in close proximity to each other, the effect of ground attenuation may not be great. A number of studies have shown that the calls of various bird and nonhuman primate species are structurally adapted to local habitats. Frog calls, however, do not appear to be influenced by habitat structure (review in Bosch and Riva, 2004). This study may be the first demonstration of the influence of habitat structure in shaping the characteristics of acoustic signals in reptiles.

Some species are able to compensate behaviorally to maintain maximal signal transmission through habitat barriers. For example, some songbirds alter the pitch of their song in different habitats (Ratcliffe and Weisman, 1985; Hill and Lein, 1987). Alternatively, if signalers cannot compensate by changing frequencies, they can increase their movement during the chorus (Hansen, 2003). Alligator signals, especially the long-distance signal, are quite monotone and the signalers remain stationary throughout the chorus. The low dominant frequencies of alligator signals may, to some extent, compensate for reduced sound transmission and the inability of alligators to behaviorally modify sound transmission by flitting about as do birds. Low frequency sound spreads more spherically relative to high frequency sound because directionality of acoustic signals depends on wavelength (Johan and LucAlain, 2005). This might explain why some songbirds turn in several directions during bouts of singing (Johan and LucAlain, 2005), but alligators remain immobile and vocalize in one direction.

Alternatively, the low frequency of alligator acoustic signals may reflect the physiological sound production system. Vocal organs of alligators are very different from those of anurans, birds and mammals (Chen *et al.*, 1985; 2003) and alligators may not be able to produce high frequency sounds. Another hypothesis is that low frequencies indicate body size and may represent "honest signals" that have evolved to impress or frighten other alligators. Further works are needed to be done in these areas.

Further studies are required to determine the mechanisms underlying the evolution of alligator communication signals. In summary, two conclusions can be drawn from this paper: (1) Chinese alligators have an elaborate acoustic communication system with both long- and short-distance signals. The former plays an important role in long range intercommunion, whereas the latter suits for short range communication; and, (2) low frequency signals may have evolved for efficient transmission in the dense vegetative habitat in which alligators breed.

Banks, M. J., and Thompson, D. J. (**1985**). "Lifetime mating success in the damselfly *Coenagrion puella*," Anim. Behav. **33**, 1175–1183.

Bosch, J., and Riva, I. D. (**2004**). "Are frog calls modulated by the environments? An analysis with anuran species from Bolivia," Can. J. Zool. **82**, 880–888.

Brandon, L. S., Ronald, J. S., and David, K. (**2003**). "Acoustic communication ranges for northern elephant seals (*Mirounga angustirostris*)," Aquat. Mamm. **29**, 202–213.

Brown, C. H. (**1989**). "The active space of blue monkey and greycheeked mangabey vocalizations," Anim. Behav. **37**, 1023–1034.

Campbell, H. H. (**1973**). "Observations on the acoustic behavior of crocodilians," Zoologica (N.Y.) **58**, 1–11.

Chen, B. H., Hua, T. M., Wu, X. B., and Wang, C. L. (**2003**). *Research on the Chinese alligator* (Shanghai Scientific and Technological Education, Shangai), pp. 18–252.

Chen, B. H., Hua, Z. H., and Li, B. H. (**1985**). *Alligator sinensis*, Hefei (Anhui Technology, Hefei), pp. 115–216.

Ding, Y. Z., Wang, X. M., He, L. J., Shao, M., Xie, W. S., Thorbjarnarson, J., and McMurry, T. S. (**2001**). "Study on the current population and habitat of the wild Chinese alligator (*Alligator sinensis*)," Biodiversity Sci. **9**, 102–108.

Douglass, H. M. (**1989**). "Song patterns of warblers at dawn and dusk," Wilson Bull. **101**, 26–35.

Duellmann, W. E. (**1970**). "The hylidae frogs of Middle America," Monograph of the Museum of Natural History, University of Kansas. Vol. **1**, pp. 1–753.

Duellmann, W. E., and Pyles, R. A. (**1983**). "Acoustic resource partitioning in anuran communities," Copeia **3**, 639–649.

Gans, C., and Maderson, P. (**1973**). "Sound production mechanisms in recent reptiles: Review and comments," Am. Zool. **13**, 1195–1203.

Garrick, L. D. (**1975**). "Structure and pattern of the roars of Chinese alligator (*Alligator sinensis*)," Herpetologica **31**, 26–31.

Garrick, L. D., and Lang, J. W. (**1977**). "Social signals and behavior of adult alligators and crocodiles," Am. Zool. **17**, 225–239.

Garrick, L. D., Lang, J. W., and Herzog, H. A. (**1978**). "Social signals of adult American alligators," B. Am. Mus. Nat. Hist. **160**, 155–192.

Gerhardt, H. C. (**2001**). "Acoustic communication in two groups of closely related treefrogs," Adv. Stud. Behav. **30**, 99–167.

Gibson, R. M., and Bradbury, J. W. (**1985**). "Sexual selection in lekking sage grouse: Phenotypic correlates of male mating success," Behav. Ecol. Sociobiol. **18**, 117–123.

Hansen, I. J. (**2003**). "Communication breakdown: Habitat influences on the communication networks and transmission of song in the black-capped chickadee (*Poecile atricapillus*)," Dissertation for bachelor degree at the University of Northern British Columbia, pp. 3–7.

Herzog, H. A., and Burghardt, G. M. (**1977**). "Vocalization in juvenile crocodilians," Z. Tierpsychol **44**, 294–304.

Hill, B. G., and Lein, M. R. (**1987**). "Function of frequency-shifted songs of black-capped chickadees," Condor **89**, 914–915.

Janne, S. K., Rauno, V. A., Johanna, M., and Silja, P. (**2004**). "Adaptive significance of synchronous chorusing in an acoustically signalling wolf spider," Proc. R. Soc. London, Ser. B **271**, 1847–1850.

Johan, J. B., and LucAlain, G. (**2005**). *The Behavior of Animals, Mechanisms, Functions, and Evolution* (Blackwell, Oxford), pp. 236–238.

Li, S., Wang, K., Wang, D., and Akamatsu, T. (**2005**). "Echolocation signals of the free-ranging Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientialis*)," J. Acoust. Soc. Am. **117**, 3288–3296.

Lind, H., Dabelsteen, T., and McGregor, P. K. (**1996**). "Female great tits can identify mates by song," Anim. Behav. **52**, 667–671.

Liu, R. S., Yu, Q., Lei, F. M., Ding, W. N., and Zhao, X. R. (**1998**). *Research of Bird's Singing* (Science, Beijing), pp. 45–52.

Marcellini, D. (**1977**). "Acoustic and visual display behavior of Gekkonid lizards," Am. Zool. **17**, 251–260.

Marler, P. (**1973**). "A comparison of vocalizations of red-tailed monkeys and blue monkeys, *Ceropithecus ascanius and C. mitis*, in Uganda," Z. Tierpsychol **33**, 223–247.

Marten, K., Quine, D. B., and Marler, P. (**1977**). "Sound transmission and its significance for animal vocalization. 2. Tropical habitats," Behav. Ecol. Sociobiol. **2**, 291–302.

Morton, E. (**1975**). "Ecological sources of selection on avian sounds," Am. Nat. **109**, 17–34.

Ratcliffe, L., and Weisman, R. G. (**1985**). "Frequency shift in the fee bee song of the blackcapped chickadee," Condor **87**, 555–556.

Robert, L. M., Michael, D. G., and Michael, K. T. (**1995**). "Chorus structure in tarbush grasshoppers: Inhibition, selective phonoresponse and signal competition," Anim. Behav. **50**, 579–594.

Roberto, S., Paolo, G., Mauro, F., and Donato, B. (**2003**). "Vocalizations and courtship intensity correlate with mounting success in marginated tortoises *Testudo marginata*," Behav. Ecol. Sociobiol. **55**, 95–102.

Ryan, M. J. (**1991**). "Sexual selection and communication in frogs: Some recent advances," Trends Ecol. Evol. **6**, 351–354.

Selvino, R. D., and Carel, T. (**2004**). "Repeated decrease in vocal repertoire size in Streptopelia doves," Anim. Behav. **67**, 549–557.

Sergio, C., Cristina, G., and Michael, J. R. (**2003**). "Call degradation in diploid and tetraploid green toads," Biol. J. Linn. Soc. **78**, 11–26.

Snedden, W. A. (**1996**). "Lifetime mating success in male sagebrush crickets: Sexual selection constrained by a virgin male mating advantage," Anim. Behav. **51**, 1119–1125.

Snowdon, C. T., and Elowson, A. M. (**1999**). "Pygmy marmosets modify call structure when paired," Ethology **105**, 893–908.

Stella, D. T., and Charles, T. S. (**2002**). "Environmental correlates of vocal communication of wild pygmy marmosets, *Cebuella pygmaea*," Anim. Behav. **63**, 847–856.

Susan, M. D, and Gerald, S. W. (**2004**). "Function of male song in the greater white-lined bat, *Saccopteryx bilineata*," Anim. Behav. **67**, 883–891.

Thorbjarnarson, J. (**1992**). *Crocodiles: An Action Plan for Their Conservation* (IUCN, Gland, Switzerland), pp. 38–39.

Tobias, M. L., Candace, B., Robert, O. H., Sam, H. H., Masha, R., and Darcy, B. K. (**2004**). "Vocal communication between male *Xenopus laevis*," Anim. Behav. **67**, 353–365.

Valentin, A., Hansjoerg, P. K., and Marc, N. (**2004**). "Seasonal patterns of singing activity vary with time of day in the Nightingale (*Luscinia megarhynchos*)," Auk **121**, 110–117.

Vliet, K. A. (**1989**). "Social displays of the American alligator (*Alligator mississippiensis*)," Am. Zool. **29**, 1019–1031.

Volker, B. D., John, K. B. F., and Peter, J. B. S. (**2005**). "The vocal behavior of mammal-eating killer whales: Communicating with costly calls," Anim. Behav. **69**, 395–405.

Wang, X. Y., Wang, D., Wu, X. B., Wang, R. P., and Wang, C. L. (**2006**). "Congregative effect of the Chinese alligator's bellowing chorus in mating season and its function in reproduction," Acta Zool. Sinica. **52**, 663–668.

Wu, J. S., and Wang, X. M. (**2004**). "Regulation of bellowing of Chinese alligators (*Alligator sinensis*) in the wild," Zool. Res. **25**, 281–286.

Young, B. A. (**1991**). "Morphological basis of growling in the king cobra, *Ophiophagus hannah*," J. Exp. Zool. **260**, 275–287.

Zhang, F., Wu, X. B., Zhu, J. L., and Zhang, S. (**2005**). "Primary research on the activity rhythm and the behavior coding of capative-bred Chinese alligator in summer and autumn," Acta Hydrobiol. Sin. **29**, 484–494.

Zhou, S. B., Qin, W. H., Jiang, H. X., Wu, X. B., Wu, L. S., Wang, H. Y., and Shao, M. (**2004**). "Vegetation diversity of wild Chinese alligator (*Alligator sinensis*) habitats," Chin. J. Appl. Ecol. **15**, 1157–1160.

Zhu, C. G. (**1957**). "Preliminary study on the life history of Chinese alligator," Acta Zool. Sinica. **9**, 132–138.

The allocation of energy to echolocation pulses produced by soprano pipistrelles (*Pipistrellus pygmaeus*) during the wingbeat cycle

Dean A. Waters[a] and Josephine G. Wong

Institute of Integrative and Comparative Biology, The Louis Compton Miall Building, University of Leeds, Leeds LS2 9JT, United Kingdom

Soprano pipistrelles exhibit considerable plasticity in both the structure and rate of echolocation call production. In search phase in the laboratory, calls are produced either as a single pulse per wingbeat cycle, or as double pulses. The amplitude of double pulses is reduced compared to the preceding single pulse. The energy flux density of either pulse of a double pulse per wingbeat was lower than a single pulse per wingbeat, and is achieved by a combination of reduction in both duration and amplitude. The combined energy of the double pulses is not significantly greater than the single pulse which precedes it. The decision to produce a double pulse may be an indication of the bat requiring additional information from the target. The production of double pulses per wingbeat may serve to achieve a higher rate of information flow for no significant increase in energetic expenditure when a possible target is detected. The proposal that echolocation during flight comes for free as a by-product of the intimate coupling among wingbeat, respiration, and echolocation is discussed. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2713717]

I. INTRODUCTION

Echolocation involves the production of high energy acoustic signals at high repetition rates. In the laboratory, bats have been recorded producing echolocation calls with amplitudes of the order of 110 dB peak-equivalent sound pressure level (peSPL) at 10 cm (Surlykke *et al.*, 1993; Waters and Jones, 1995), while in the field, these calls may be as loud as 133 dB peSPL (Holderied and von Helversen, 2003; Holderied *et al.*, 2005). Given the low efficiency of conversion of metabolic energy to acoustic energy, typically between 0.05% and 6% (Prestwich, 1994; Prestwich and O'Sullivan, 2005) calls of these intensities should require large amounts of energy to produce. Yet, by coupling echolocation with the muscles used for flight, bats essentially get echolocation "for free" (Speakman and Racey, 1991). Many studies have shown that there is a tight correlation between the wingbeat cycle and echolocation call production for bats which echolocate in flight (Herbert, 1985; Kalko, 1994; Britton *et al.*, 1997) and it is noticeable that bats produce lower intensity echolocation calls when stationary (Waters and Jones, 1995). The exception to this are bats which perch hunt which rely less on muscular activity, and more on the elastic properties of the rib cage and abdominal aponeurosis to efficiently echolocate from a stationary perch (Lancaster *et al.*, 1995; Lancaster and Speakman, 2001).

The normal relationship of one call per wingbeat cycle does not hold when bats have detected a target and change from the search phase to the approach phase and the terminal buzz, where repetition rates may reach 200 Hz (Kalko, 1995). Some species of bats also produce extra pulses during a single wingbeat cycle during normal search phase (Kalko, 1994). These extra pulses have been termed "double pulses" (Wong and Waters, 2001) or "strobe groups" (Moss *et al.*, 2006). In *Pipistrellus pygmaeus* (Leach), single echolocation calls are produced either before the top of the upstroke or just after (Kalko, 1994; Wong and Waters, 2001). When double pulses are produced, the bat produces an echolocation call at each location in the wingbeat cycle (Wong and Waters, 2001). During the transition to the approach phase, echolocation calls apparently increase in bandwidth and repetition rate and reduce in duration (Kalko, 1995). There is also a reduction in amplitude as the bat approaches a target (Boonman and Jones, 2002). Jen and Kamada (1982) suggested that this response was to either keep returning echoes within an optimal level, or because of physiological limitations at simultaneously increasing repetition rate and maintaining output amplitude.

Here, we investigate both the structure and intensity of double pulse production in *P. pygamaeus* to explore the link between repetition rate and acoustic energy output. We train bats to fly on a stereotypic flight path along a flight tunnel. A target is present on a randomly allocated basis to encourage the bats to actively probe their surroundings. We investigate the characteristics of consecutive echolocation calls produced as single pulses per wingbeat, consecutive double pulses per wingbeat, and a single pulse per wingbeat followed by a double pulse per wingbeat. In particular, we compare the energy flux density of pulses in order to investigate the potential costs of pulse production.

---

[a]Electronic mail: d.a.waters@leeds.ac.uk
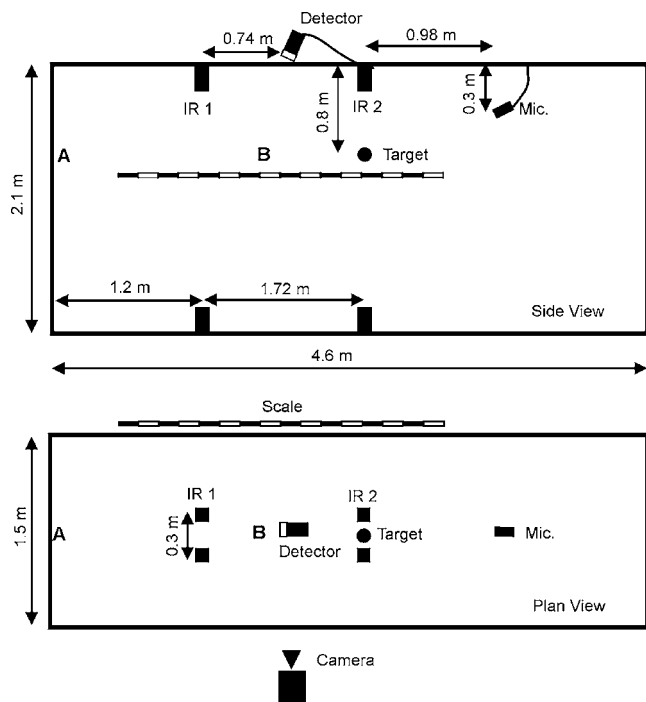
FIG. 1. Schematic diagram of the flight tunnel. (A) The position of the bat when released, and (B) the approximate location of the three call sequence used for analysis. IR1 and IR2 are the two pairs of infrared triggers. The detector shown between IR1 and IR2 is used to synchronize vocalizations with the high speed video, and the recording microphone is shown behind the target location. The camera location is not to scale.

II. MATERIALS AND METHODS

A. Subjects

Six soprano pipistrelles (*Pipistrellus pygmaeus*) (identified as PP1–PP6) were kept in the laboratory on a staple diet of mealworms (*Tenebrio molitor*) supplemented with buffalo worms (*Alphiobius diaperinus*), waxmoths (*Galleria mellonella*), black crickets (*Gryllus bimaculatus*), and multivitamins. Bats were maintained at a weight of between 5 and 6 g and allowed free flight in a $2.5 \times 3 \times 2.5$ m room with a programmed light:dark cycle (12 h:12 h) set such that experiments started in the morning, 1 h after the beginning of the night cycle.

B. Bat's task

Bats were trained to fly on a straight flight path from one end to the other of a $4.6 \times 2.1 \times 1.5$ m flight tunnel (Fig. 1) constructed from steel frame and plastic netting (mesh size 1 cm). The flight tunnel was placed in a larger room (6 $\times 2.6 \times 7$ m). Two sets of infrared triggers were placed along the flight path configured such that if the bat broke either of the beams in each pair, a reference transistor-transistor logic (TTL) pulse was recorded. The pairs of beams were positioned 30 cm apart with 1.72 m between the pairs along the length of the tunnel (Fig. 1). Recordings were accepted only if one or both beams from both pairs of infrared triggers were broken ensuring that the flight was straight and along the main axis of the flight tunnel. The bat flew toward the location of a target consisting of a 60-mm-diam, 3-mm-thick black perspex disk, suspended by a 0.15-mm-diam nylon

monofilament line situated 1.68 m from the back wall of the flight tunnel between and below the second set of infrared beams. The target was either present or absent per run on a pseudorandom basis such that the target was present on half of the runs. This was to discourage familiarity with the tunnel layout and to ensure the bat actively echolocated to probe its surroundings. Only data from runs where the target was present were used in the full analysis.

Echolocation calls were recorded from a U30 bat detector (solid dielectric capacitance type ±3 dB, 20–120 kHz, Ultrasound Advice, London, UK), suspended 98 cm behind the target location and 50 cm above it, sampled by a DAS1800AO data acquisition card (Keithley, Theal, Berkshire, UK) at 391 kHz and 12 bit resolution. Bats were filmed side-on using a high-speed SR-500 Kodak Motion Corder Analyzer (Eastman Kodak, New York) at a rate of 250 frames s^{-1}, resulting in a time resolution of 4 ms between images. We used a Pentax Cosmicar fl.8 20 mm zoom (Pentax, Japan) fitted with a $\times 0.42$ wide angle lens attachment. The camera was placed outside the flight tunnel so that the optical axis was normal to the long axis of the flight tunnel at a height of 1.3 m. The distance from the lens to the edge of the long axis of the flight tunnel was 2.17 m. In order to calculate the distance of the bat to the target or target position, a horizontal scale parallel to the long axis of the flight tunnel was marked out 0.2 m behind the flight tunnel. The scale was marked in black on a white background so that it remained visible through the flight tunnel netting. Runs were accepted only if the bat flew parallel to this scale. To control for the effects of curvature caused by the wide-angle lens, the horizontal scale was calibrated by first marking increments of 0.15 m from the first pair of infrared sensors to the target in the center of the flight tunnel using tape markers on a wire. These temporary markers were then viewed through the video camera, and the corresponding 0.15 m increments (as seen through the camera) were transcribed onto the scale behind the flight tunnel allowing estimation of the position of the bat along the flight tunnel to within 3.5 cm depending on which quarter of each scale increment the bat's head was located closest to. Since the bat was not filmed from in front, as in Wong and Waters (2001), it was not possible to establish the exact angle of the wings at the time of pulse emission. It was however possible to monitor whether the wings were in the upstroke or downstroke phase and the approximate part of the phase when the pulses were emitted since a single wingbeat cycle covered approximately 16 frames of video, providing theoretically 22.5° of resolution in wing angle. Since it was not possible to estimate the wing angle to this level of accuracy viewed from the side, echolocation pulses were assigned to one of eight positions. These were: two locations between the highest elevation at the beginning of the downstroke and the wing-horizontal position; two below horizontal to the maximum depression of the wings at the bottom of the downstroke; two from this position to the wing-horizontal position on the upstroke; and two between horizontal to the maximum elevation on the upstroke.

The captured images were transferred at a playback rate of 5 frames s^{-1} to Hi-8 tapes using a Sony GV-A5000E video Walkman (Sony, Japan).

Collection of both video and echolocation data was triggered automatically when the bat broke one of the first set of infrared triggers and was synchronized together by taking the heterodyne output from a Batbox III bat detector (Stag Electronics, Steyning, West Sussex, UK) suspended between the two sets of infrared beams into one channel of a Kodak multichannel data link. This resulted in a visible indication on the video frame when the bat echolocated. The maximum time discrepancy between the reception of an echolocation call and the synchronization with the video frame due to the speed of sound was 3.2 ms, resulting in a maximum error of 1 frame, with a typical full wingbeat occupying 16 frames. Additional channels on the multichannel data link recorded data from the infrared triggers indicating whether the bat was flying down the center of the flight tunnel. With this arrangement, it was possible to know whether the bat was flying directly toward the target and recording microphone, how far away from the target location and microphone each bat was when it echolocated, and to collect those same echolocation calls for later analysis.

Each bat was released 1.2 m behind the first pair of infrared beam sensors at a height of about 1.3 m and flew toward the target. Correct flights in which the bat did not land or swerve off a straight course were rewarded with a waxmoth or a buffalo worm once the bat had landed on the side or end of the tunnel. Each bat completed ten trials per day until a total of forty trials were completed from each of the six bats resulting in a total of 240 experimental runs. All experiments were performed under dim red light and an infrared LED source (Tracksys Ltd., Nottingham, UK) resulting in an overall illumination of 0.14 mW m^{-2} (J16 digital photometer, Tektronix, Beaverton, Oregon).

III. DATA ANALYSIS

Video recordings were replayed on a 21 in. KV21X5 Sony monitor using a Sony GV-A5000E video Walkman (Sony, Japan). The distance from the bat to the disk during runs with a real target and the distance from the bat to the position of the target during no target runs (target range) were calculated from the background scale for each echolocation pulse. From this, the distance to the recording microphone could be calculated. Echolocation call sequences were analyzed using a custom program written in TESTPOINT (Keithley, Theal, Berkshire, UK). Measurements of the pulse interval (ms) and the pulse duration (ms) of echolocation pulses from the wave forms were taken to an accuracy of 0.05 ms. The pulse interval was taken as the time interval between the start of the pulse to the start of the next pulse. The minimum, maximum, and peak frequency of the fundamental was measured using a 1024-point Fast fourier transform (FFT) power spectrum with Hamming window. The bandwidth of the signal was calculated by subtracting the minimum frequency from the maximum frequency at 30 dB below the peak frequency.

Three consecutive calls produced as single pulses per wingbeat (here termed SP1, SP2, and SP3) were selected from runs where the target was present for five bats from between 4 and 10 runs per bat depending on whether the bat produced the required pulse combinations at the required target distance within the video field of view. In addition, the above-mentioned measurements were also made for consecutive pulses emitted as double pulses per wingbeat (termed DP1a and DP1b, followed by DP2a and DP2b taken from 4 to 10 runs per bat), and for a single pulse followed by a double pulse per wingbeat (SP1 followed by DP1a and DP1b taken from between 9 and 10 runs per bat). For consistency in the analysis, only three of the pulses were used in the sets where two double pulses were produced in a row; the second of the pulse of the first set of double pulses (DP1b) followed by DP2a and DP2b. Each sequence of different call types was taken from a different run. One individual bat was omitted from the analysis due to the lack of double pulse per wingbeat production. Calls were chosen from approximately the same distance to the target in each run corresponding roughly to 1 m from the target location (location B in Fig. 1) and 2 m from the recording microphone.

Peak amplitude and the (rms) pressure of signals was measured in Adobe Audition (Adobe Systems Inc., San Jose, CA) and later converted to actual sound pressures by calibrating the U30 microphone using a 55 kHz sine wave from an Ultrasound Advice loudspeaker and a Larson Davis 1/4 in. 2520 microphone, 910B preamplifier and 2200C measuring amplifier (Larson Davis Inc., Provo, UT) with the protective grid off at a distance of 2 m. The Larson Davis microphone was calibrated with reference to a Dawe D-1411E acoustical calibrator (Lucas CEL Instruments, Hitchin, UK). Assuming the inverse square law for the intensity-distance relationship, and therefore the inverse distance law for the pressure-distance relationship (i.e., a 6.02 dB decrease in intensity with doubled distance), calls amplitudes were corrected for different distances of individual signals from the recording microphone since the distance from the bat to the target location was known and the distance from the target to the recording microphone was constant and known. Sound pressure levels were then standardized to 0.1 m. Excess atmospheric attenuation would have a small additional effect over the small range of distances encountered when looking within a three call sequence [<0.5 dB at 55 kHz (Bazley, 1976)], but no initial compensation was made for this in the raw data since the total amount of energy loss depends on the frequency spectrum, and unlike spherical spreading loss, is difficult to estimate accurately for a broadband signal. Energy flux density (J m^{-2}) and crest factors of pulses were calculated using equations from Au (1993) and Waters and Jones (1995) for standardized distances of 1 m. Energy flux density is a measure of the energy content of a call, and therefore provides information on the metabolic costs of production since peak amplitude simply reflects the maximum intensity of the signal, rather than the amount of time spent at peak intensity. Crest factors are indicative of the time at peak amplitude and

D. A. Waters and J. G. Wong: Energy in echolocation pulses of bats
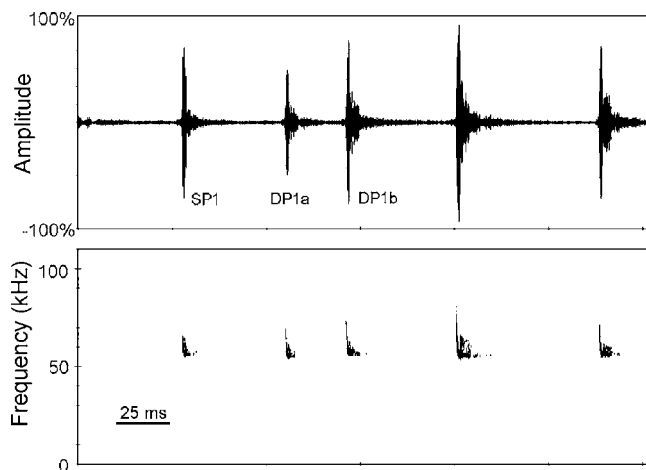
FIG. 2. Wave form and sonogram of a sequence of calls from within the tunnel showing a five call sequence consisting of a single pulse (SP1) followed by a double pulse (DP1a and DP1b) followed by two more single pulses.

were calculated by dividing the peak amplitude by the rms of a pulse and are useful indicator of the envelope shape of the pulse.

A one-factor repeated measures or within subjects ANOVA was performed (Zar, 1999) to compare the parameters of three calls produced sequentially; three single pulses in a row, the second call of a double pulse pair with the two calls in the subsequent double pulse, and a single pulse followed by a double pulse. Parametric statistics were employed as the data did not depart from the underlying assumption of normality. Following significant main effects of the within subjects factors, the Bonferroni method was adopted as a multiple comparison procedure following Kinnear and Gray (1999). Thus, pairwise comparisons were made using paired-sample t-tests between the means for combinations of the call types. To achieve significance in the Bonferroni method, paired t-tests must show significance beyond the $P=0.025$ level where two paired comparisons are made, and beyond the $P=0.017$ level where three paired comparisons are made. Typical multiple comparison procedures such as the Tukey or Neuman-Keuls test were not employed due to the increased probability of Type I errors if the Mauchly's test of sphericity for homogeneity of covariance was significant and the sphericity assumption violated. In the present paper, in cases where the sphericity assumption was violated in the repeated measures ANOVAs, the more conservative Greenhouse-Geisser test was used. All analyses were two-tailed and run on SPSS, version. 13 (SPSS Inc., Chicago, IL).

IV. RESULTS

A sequence where a bat produces a single pulse followed by a double pulse is shown in Fig. 2. Plots of pulse interval with distance to the target (pooled across runs within an individual) clearly show an overall bimodal distribution for both where the target is present [Fig. 3(a)] and absent [Fig. 3(b)]. Such a distribution is the result of the bat mixing the production of single pulses per wingbeat and double pulses per wingbeat. The alternation between the two pulse types is
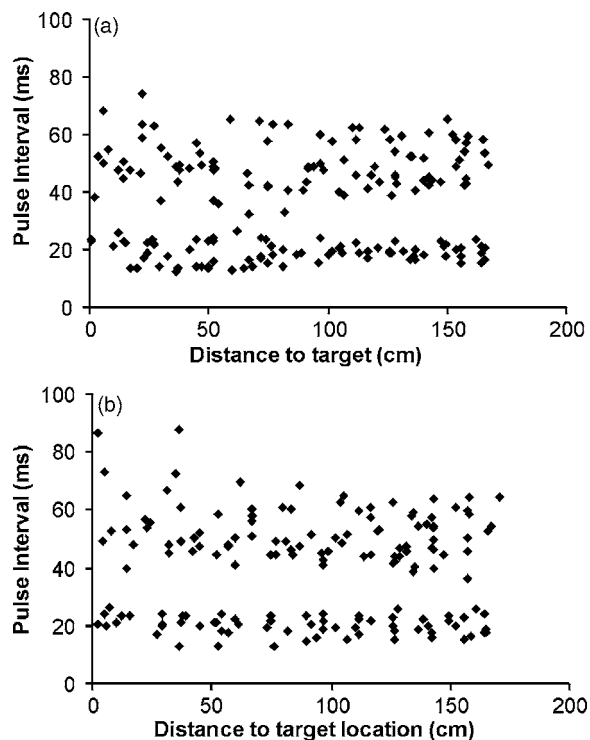
FIG. 3. Pulse interval as a function of target range for PP3 under (a) target present and (b) target absent conditions ($N=10$ pulse trains for each target condition). The two classes of pulse intervals can be seen, those above 40 ms and those around 20 ms.

more clearly seen in Fig. 4, where the results from a single run are presented, the bat producing single pulses per wingbeat (here termed SP), and hence long interpulse intervals, followed by two pulses per wingbeat (here termed DP) with short interpulse intervals. This alternation tends to be produced in a 1:1 ratio when the target is present, and 1:2 ratio when the target is absent, i.e., two single pulses per wingbeat followed by a double pulse per wingbeat.

While the exact angles of the wings at which echolocation pulses were produced could not be established since the bats were filmed from the side, the approximate points of pulse production were just prior to the top of the upstroke and just after the start of the downstroke. No pulses were observed where the wings were below horizontal. Single pulses were mainly produced prior to the top of the upstroke (72% of 36 experimental runs across five bats), while double pulses were produced by combining a call prior to the top of the upstroke with another just after the start of the downstroke (85 experimental runs across five bats).

Comparisons were made between three consecutive calls at the approximate midpoint of the bats' flight for three situations where the target was present: Three single calls per wingbeat; a single pulse followed by a double pulse; and the second pulse of a double pulse followed by the two calls of the second double pulse in a row. Call intensity, pulse duration, energy flux density, crest factor, and bandwidth of the fundamental frequency did not differ significantly between SP1 and SP2 or between SP2 and SP3 in the three consecutive single pulses per wingbeat [Figs. 5(a)–5(c)] indicating that signal structure did not differ between three consecutive single pulses at that point in the flight tunnel. Mean call
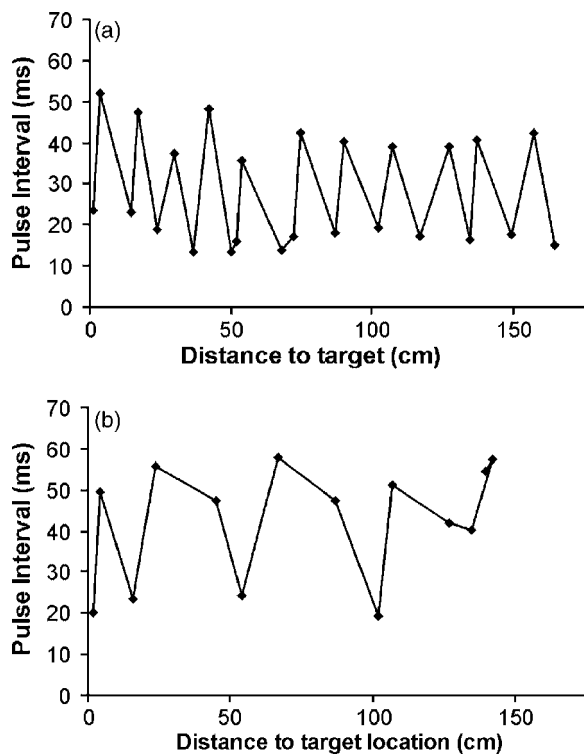
FIG. 4. Pulse interval as a function of target range for PP3 under (a) target present and (b) target absent conditions ($N=1$ pulse train for each target condition). The alternation in pulse intervals between long and short is apparent, with a 1:1 ratio when the target is present, and a 2:1 ratio when it is absent. In (a), the bat alternates between long pulse intervals and short pulse intervals corresponding to a double pulse. In (b), the bat uses two long pulse intervals in a row followed by a short pulse interval.

intensity was 106 ± 3.4 dB peSPL at 10 cm, duration was 1.8 ± 0.29 ms, energy flux density $3.3 \pm 2.30 \times 10^{-7}$ Jm^{-2}, crest factor 5.5 ± 0.14, and bandwidth 53.1 ± 3.8 kHz. Furthermore, these parameters did not differ when considering the second pulse of the double pulse pair (DP1b) compared with the first pulse of the next consecutive double pulse pair (DP2a) or between DP2a and the second pulse of the same double pulse pair (DP2b) in the three consecutive pulses [Figs. 6(a)–6(c)]. In this situation, mean call intensity was 101 ± 1.8 dB peSPL at 10 cm, duration was 1.1 ± 0.28 ms, energy flux density $6.2 \pm 2.20 \times 10^{-8}$ Jm^{-2}, crest factor 5.3 ± 0.47, and bandwidth 46.3 ± 2.79 kHz. Together, these findings rule out any systematic longitudinal effect of range or position of the bat in the flight tunnel on the three calls under investigation. It is notable however that call intensity, and hence energy, as well as duration and bandwidth, were lower for the three double pulses in a row than the three single pulses in a row, when compared between runs.

Within runs, mean call intensity was significantly lower for the first pulse of a double pulse pair (DP1a) than for the preceding single pulse SP1 (paired t-test, $t=3.99$, $df=4$, $P < 0.025$) [Fig. 7(a)], with the mean intensity for DP1a being 102.4 ± 2.23 dB peSPL compared with 106.4 ± 2.23 dB peSPL for SP1. This difference was on average 4 dB. No significant difference was found between the mean call intensity of the second pulse of a double pulse pair (DP1b) and the preceding single pulse (paired t-test, $t=1.43$, $df=4$, $P > 0.05$) with the mean call intensity of DP1b being
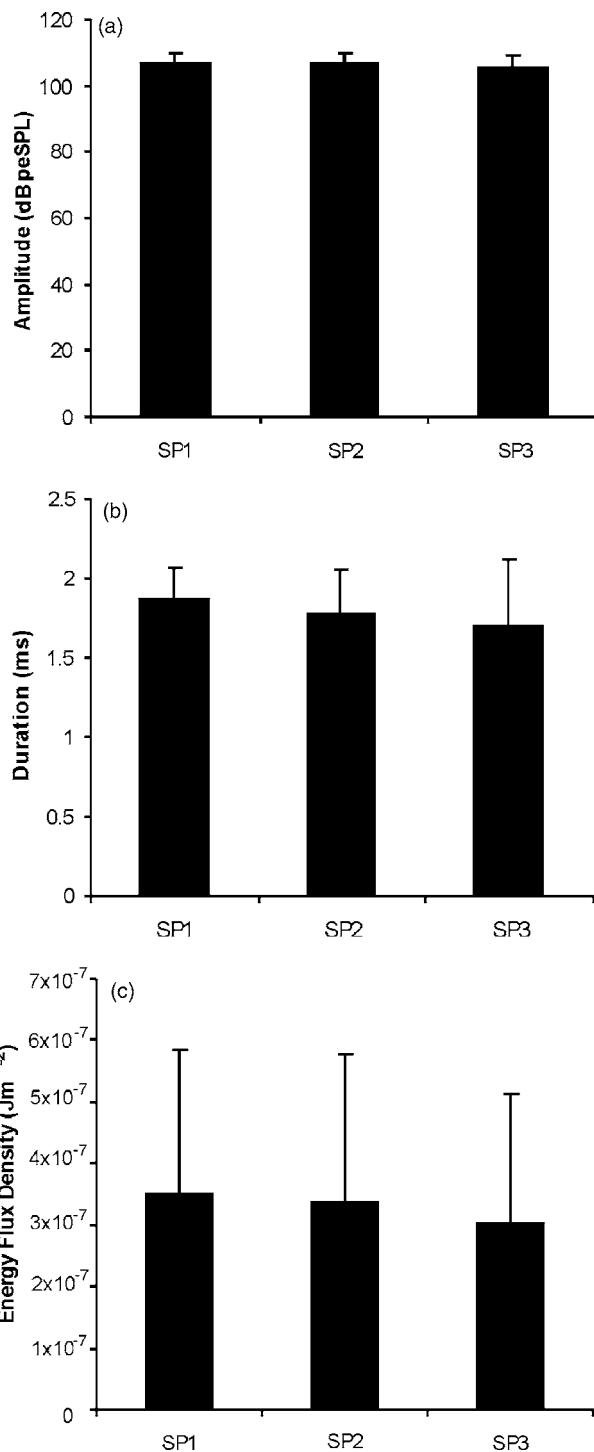
FIG. 5. Mean call intensity (a), pulse duration (b), and energy flux density (c) for three consecutive single pulses per wingbeat obtained from five bats ($N=4-10$ consecutive single pulses per individual). Values are mean±s.d.

105.4 ± 3.1 dB peSPL. There was a significant difference in pulse duration among SP1, DP1a, and DP1b [repeated measures ANOVA (Greenhouse-Geisser), $F_{(1,6)}=8.5$, $P < 0.05$]. However, it was not possible to determine if single pulses differ significantly in pulse durations than either pulse of a double pulse pair (Bonferroni corrected paired t-test; single pulse versus DP1a: $t=2.83$, $df=4$, exact $P=0.047$, single pulse versus DP1b: $t=2.92$, $df=4$, exact $P=0.043$ where Bonferroni corrected critical $P=0.025$). There is an indica-
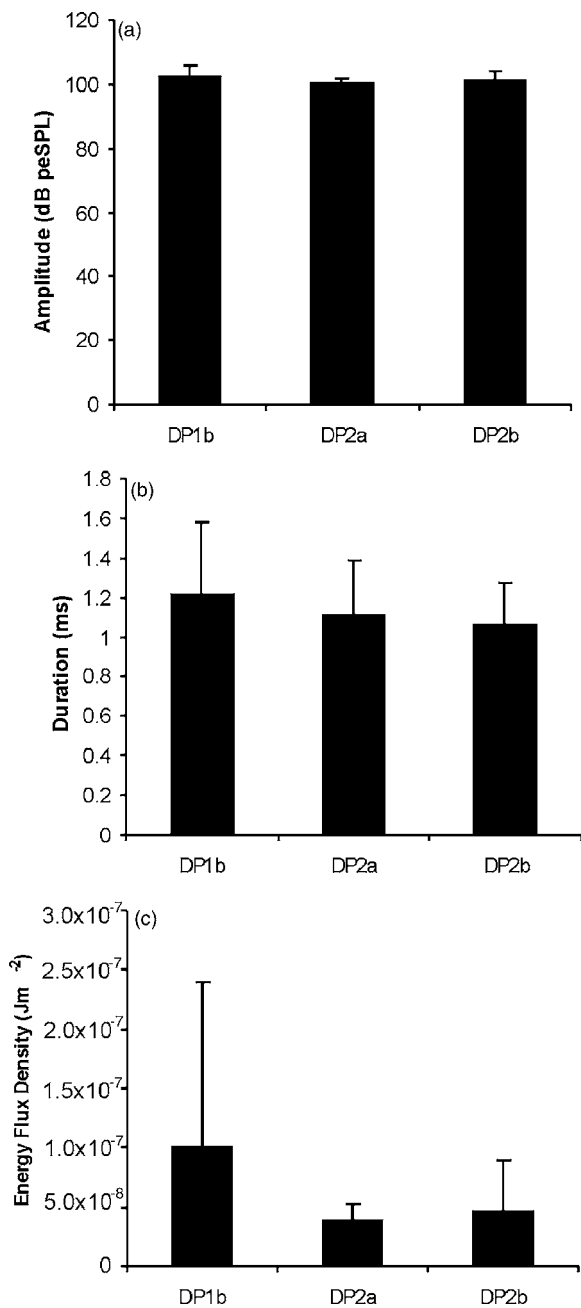
FIG. 6. Mean call intensity (a), pulse duration (b), and energy flux density (c) for three consecutive pulses emitted as double pulses per wingbeat from five bats ($N=4-10$ consecutive double pulses per individual). DP1b represents the second pulse of the first double pulse pair, DP2a and DP2b represent the first and second pulse, respectively, of the next consecutive double pair. Values are mean±s.d.

FIG. 7. Mean call intensity (a), pulse duration (b), and energy flux density (c) for a single pulse per wingbeat followed by a double pulse per wingbeat from five bats ($N=9-10$ consecutive single pulses and double pulses per individual). DP1a and DP1b are the first and second pulse, respectively, of the double pair following the single pulse per wingbeat, SP1. Values are mean±s.d.

tion from Fig. 7(b) that both DP1a (1.51±0.40 ms) and DP1b (1.55±0.37 ms) have reduced durations compared with SP1 (1.81±0.26), so while not significant after Bonferroni correction, the ANOVA is most likely indicating a difference in duration between SP1 and both DP1a and DP1b. There were also significant differences in the energy flux density between SP1 ($4.4±2.75×10^{-7}$ J m^{-2}), DP1a ($1.8±1.59×10^{-7}$ J m^2) and DP1b ($3.5±2.59×10^{-7}$ J m^{-2}), [repeated measures ANOVA (sphericity assumed), $F_{(2,6)}=10.9$, $P<0.01$, Fig. 7(c)]. However, Bonferroni corrected paired t-tests failed to reveal where the differences occurred

(paired t-test; single pulse versus DP1a: $t=3.20$, $df=4$, exact $P=0.033$); single pulse versus DP1b: $t=2.91$, $df=4$, exact $P=0.043$ where Bonferroni corrected critical $P=0.025$). The greatest difference is between SP1 and DP1a where, on average, DP1a contained 40% of the energy of SP1 which preceded it, while DP1b contained on average 79% of the energy of SP1. The sum of energy flux density for DP1a and DP1b within a run was not significantly greater than that of

the single pulse that preceded the double pulse pair (paired t-test; $t=1.09$, $df=4$, $P>0.05$), with the double pulses together amounting to 119% of the energy in the preceding single pulse. Pairwise comparisons revealed no significant differences in mean crest factors (mean 5.8 ± 0.21) or in bandwidths (mean 50.4 ± 5.60 kHz) between single pulses and either pulse of the double pulse pair meaning that the frequencies and envelope shapes were similar for SP1 and DP1a and DP1b.

V. DISCUSSIONS

This study has compared three different types of pulse production produced under identical conditions in a flight tunnel; the production of three consecutive single pulses, two consecutive pairs of double pulses, and a single pulse followed by a double pulse. In all cases, a sequence of three individual echolocation calls, produced at approximately the same location in the tunnel, with amplitudes corrected for distance to the recording microphone are compared.

The production of double pulses per wingbeat per run resulted in a characteristic alternation in interpulse intervals (Figs. 3 and 4) where bats generally produced an alternation between single pulses and double-pulse pairs with an apparent increase in double pulse production when the target was present. Wong and Waters (2001) showed that in *P. pygmaeus*, single pulses could be produced either shortly before the top of the upstroke, or just after the beginning of the downstroke, individual bats showing a preference for one mode or the other. This finding is confirmed in this current study, where most single pulses were produced just prior to the top of the upstroke, with a minority just after the start of the downstroke. Double pulses were produced at the same wingbeat cycle positions as single pulses, but were a consequence of the bat echolocating at both locations within one full wingbeat cycle resulting in a shortened interpulse interval. A similar echolocation pattern has been observed in a number of other species. Grinnell and Griffin (1958) described the regular alternation of pulse intervals between two different values in *Myotis lucifugus* at about 1 m away from wire obstacles, a pattern also seen in *E. fuscus* (Jen and Kamada, 1982), *A. tridens* (Gustafson and Schnitzler, 1979), *Desmodus rotundus* (Joermann and Schmidt, 1981), and *P. parnellii* during obstacle or target approach (Lancaster *et al.*, 1992). Moss *et al.* (2006) documents these clusters of echolocation calls seen in *Eptesicus fuscus*, and describes them as "strobe groups," most often seen when the bat encounters a target.

The experiments discussed here were conducted within the confines of a flight tunnel which may influence the rate of double pulse production. Double pulses are not apparent in data shown for *P. pygmaeus* flying in a larger flight tunnel (Surlykke *et al.*, 2003), but can be seen during the approach phase of prey pursuit in the field in *P. pygmaeus* and *M. brandtii* [Author's data, Figs. 8(a) and 8(b)], *M. daubentonii* (Wong, unpublished data), and *M. capaccinii* [from E. Kalko (unpublished data) cited in Neuweiler (2000)]. Kalko (1994) also reports that for pipistrelles, recordings of the search phase "often showed a pattern of loud signals interspersed
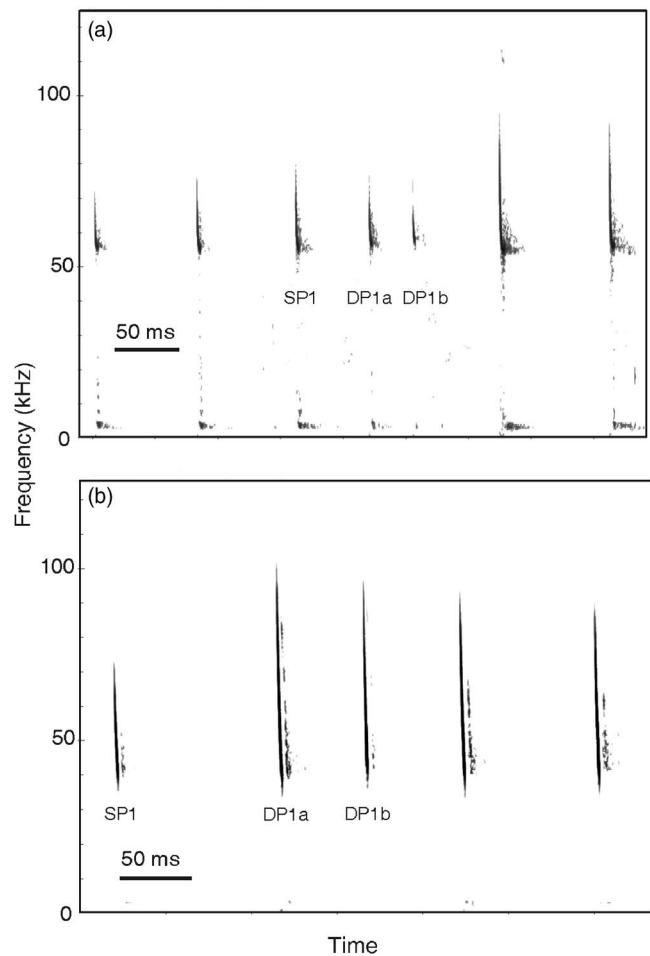
FIG. 8. Sonograms of echolocation calls produced during search flight in the field by individuals of (a) *P. pygmaeus* and (b) *Myotis brandtii* showing double pulse production. Note that since the angle and distance of the bat to the microphone is unknown, temporal relationships between the call types are accurate but amplitude relationships between the call types cannot be directly compared. SP1 is a single pulse per wingbeat, DP1a and DP1b are the first and second pulse of a double pulse per wingbeat.

with signals of lesser intensity." While the rate of double pulse production does appear to increase when the bat is flying in a small tunnel, and when an additional target is present, double pulses do appear to be a real phenomenon when the bat is flying under normal field conditions.

Wong and Waters (2001) suggest that when bats increase repetition rate by emitting two pulses per wingbeat, pulse emission remains highly synchronized with the wingbeat and respiratory cycle. Thus, the energetic cost of producing a double pulse per wingbeat may be similar to producing a single pulse per wingbeat if timing is the only issue. The energy flux density of pulses was calculated in order to investigate the energy output of sound pulses [see Au (1993) and Waters and Jones (1995) for a full description of rationale and method]. Although we found no difference in the mean energy flux density between three consecutive single pulses per wingbeat and also no difference between consecutive pulses emitted as double pulses per wingbeat, the energy content of both the first and, and possibly, second pulse of a double pulse pair was lower than the preceding single pulse [Fig. 7(c)].

Energy flux density is directly related to call amplitude and duration and inversely related to crest factor (Waters and Jones, 1995). Call intensity was significantly reduced for the first pulse of a double pulse pair when compared with the preceding single pulse by an average of 4 dB [Fig. 7(a)]. When bats produced a double pulse pair, pulse duration was also reduced when compared with a single pulse [Fig. 7(b)]. In contrast, no significant difference was found in pulse duration either between three consecutive single pulses [Fig. 5(b)] or between three consecutive pulses emitted as double pulses per wingbeat [Fig. 6(b)]. Waters and Jones (1995) calculated crest factors by dividing the peak amplitude by the rms amplitude of a pulse. Therefore, a higher crest factor indicates a relatively shorter amount of time at peak amplitude. No significant difference was found in crest factors between a single pulse and either pulse of the double pulse pair, which suggests that the shape of the wave form of a single pulse and the first and second pulse of a double pulse pair were relatively similar. We therefore conclude that the higher energy flux density of single pulses was probably the result of the combination of longer call duration and higher call amplitude. In contrast, the relatively lower energy flux density of either call in a double pulse pair was due to the combination of shorter call durations and lower call amplitudes. We did not control for the loss of sound energy through excess atmospheric attenuation since it depends on the overall frequency range of the signal. However, the overall loss atmospheric attenuation over the distance a bat flies over a three call sequence (less than 0.5 m) is small when compared with the magnitude of any observed differences in call amplitude, and any compensation for atmospheric attenuation would also accentuate the observed differences rather than reduce them [e.g., Figs. 7(a) and 7(c)].

The energy reduction is especially marked in the first double pulse following a single pulse with the mean energy flux density being only 40% of the single pulse that precedes it. The total energy flux density of a double pulse pair is about 119% that of a single pulse. The pattern that *P. pygmaeus* uses is a normal single pulse, followed by a reduced energy pulse, quickly followed by a second, slightly reduced energy pulse.

Potentially, by producing double pulse pairs, bats could gain twice as much information per unit time for only a little more extra energetic cost over that used to produce a single pulse. However, this increase in information flow, achieved through the production of double pulse pairs, probably only occurs once the bat has detected and "locked-on" to a potential target as shown by the increased production of double pulse pairs when a target is present (Fig. 4).

The use of double pulses per wingbeat, with their lower energy content, should result in lower amplitude echoes and presumably a reduced maximum range of detection. A reported reduction in amplitude of 6 dB per halved distance would result in constant amplitude at the target, but a perceived increase in echo intensity of 6 dB per halving of range (Hartley, 1992a). Such an increase in echo amplitude would compensate for the reduction in hearing sensitivity of 6−7 dB per halving of range caused by automatic gain control (Hartley, 1992b), though Simmons *et al.* (1992) reports

this figure to be 11 dB per halving of range, close to the 12 dB necessary to offset the increase in echo intensity if the bat did not reduce its call amplitude. While it would seem desirable to maximize the SPL of a returning echo, Denzinger and Schnitzler (1994) found that echoes that return at −10 dB or louder relative to the output SPL of the call show reduced ranging capabilities. This was proposed to be consequence of the paradoxical latency shift whereby neurons of the auditory cortex show shorter latencies when sounds are of low amplitude compared to high (Sullivan, 1982). A reduction in amplitude may also be a consequence of energy conservation or limitation in that it may simply not be possible to produce sufficient calls at full amplitude at high repetition rates (Jen and Kamada, 1982). However, the lower amplitudes of echoes from double pulses is a consequence of their lower output amplitudes, maintaining a constant output/echo energy relationship and does not explain why a double pulse is then followed by a louder single pulse.

The reduced interpulse interval of double pulses would limit bats to specific target ranges to avoid the coincidence of echoes from the first pulse while producing the second. From the data derived from these experiments for interpulse intervals, this would limit the use of double pulses to target ranges of 3.2 m or less. Interestingly, because the energy content of the first (as well as the second) pulse of a double pulse is lowered, this suggests that the bat may decide whether or not to produce a double pulse following the preceding single pulse rather than produce a loud single pulse immediately followed by a quieter pulse with a shorter interpulse interval. Such an arrangement could imply that the bat is probing a target with a double pulse following initial detection. An alternative explanation is that the double pulse production follows a preprogrammed emission pattern that the bat does not dynamically change. However, given the variability of response in terms of double pulse production both between bats and within bats between runs, this explanation does seem less likely.

During the production of a double pulse per wingbeat cycle, the two pulses will typically be separated by less than 30 ms and are produced at two locations; just before the top of the upstroke and just after the start of the downstroke. In the laboratory however, *P. pygmaeus* shows a preference for echolocating with single pulses at either location, while the locations for the production of the double pulses are fixed. If the individual bat is echolocating with single pulses at the start of the downstroke, then the double pulses will be produced at the top of the upstroke and start of the downstroke, resulting in a short time between SP1 and DP1a [as in Fig. 8(a) and shown schematically in Fig. 9(a)]. However, if the bat is emitting single pulses just prior to the top of the upstroke, then there will be a longer interval between SP1 and DP1a as the first of the double pulses is not produced until the wings are at the same position as the preceding single pulse as shown schematically by Fig. 9(b).

Not all bat species produce echolocation calls at the same location in the wingbeat cycle. Suthers *et al.* (1972) reported a complex pattern in *Phyllostomus hastatus* depending on the number of pulses produced, the usual pattern being a pulse produced at the end of the inspiration/start of the
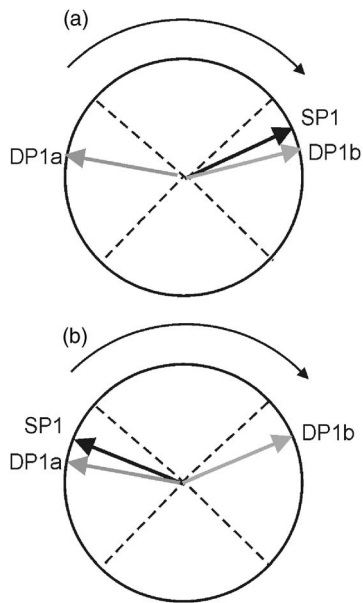
FIG. 9. Locations of pulse production during the wingbeat cycle of two alternative schemes for a single pulse (SP1) followed by two double pulses (DP1a and DP1b). Dotted lines represent the maximum extent of the upstroke and downstroke. The scheme in (a) shows SP1 being produced at the start of the downstroke, with DP1a following at the top of the upstroke, followed by DP1b at the start of the next downstroke. This results in a short interval between SP1 and DP1a. In (b), SP1 is produced at the end of the upstroke, with DP1a not being produced until the same position is reached again, then followed by DP1b at the start of the downstroke. This results in a longer interval between SP1 and DP1a.

expiration when the wings were at the downward-most part of the wingbeat phase. At higher repetition rates a pulse will be produced at the start of expiration (downward-most) and the start of inspiration (upward-most). Joermann and Schmidt (1981) found that *Desmodus rotundus* emitted echolocation calls at the start of the upstroke and Moss *et al.* (2006) found that *Eptesicus fuscus* produces clusters of calls (termed strobe groups), which were not closely correlated to a particular phase of the wingbeat cycle. These strobe groups were produced more often in cluttered conditions and on approach to the target and consisted of groups of echolocation calls separated from other groups by longer pulse intervals. Strobe groups appear analogous to the double pulses reported here but with less tight correlation with the wingbeat cycle. It is argued that these strobe groups sharpen spatial response fields in the bats midbrain and allow summation of spatial information over time (Moss and Sinha, 2003 cited in Moss *et al.*, 2006) rather than simply providing more rapid single "snapshots" of spatial information from single pulses.

Echolocation calls are not totally phase-locked to the wingbeat cycle, as evidenced by the high repetition rates during the terminal buzz, but do show a general pattern of linkage to the expiration cycle that occurs during the upstroke (Lancaster *et al.*, 1995). While the role of the pectoralis and serratus ventralis muscles in powering the upstroke and assisting the lateral abdominal wall muscles to pressurize the thoracic cavity has been demonstrated (Lancaster *et al.*, 1995), it is the abdominal muscles alone that show an obligate relationship with echolocation. Subglottic pressures rise to 40 cm H_2O prior to echolocation, and drops between

5 and 20 cm H_2O during the course of each pulse with a close correlation between subglottic pressure and echolocation intensity (Fattu and Suthers, 1981). The maximum subglottic pressure, and hence the maximum echolocation call intensity, is probably limited by the necessity of maintaining the blood flow through the lung capillaries since high backpressure from the thoracic pressurization would interrupt oxygenated blood-flow from the lungs back to the heart (Suthers, 1988). When calls are produced in rapid succession, they are produced during a single exhalation, the subglottal pressure dropping after each pulse, then recovering quickly. A different approach is used by horseshoe bats which take a "minibreath" after each echolocation call (Suthers, 2004), but adaptations of the abdominal musculature, and the fact that these bats perch hunt, may reduce the link between flight, breathing and echolocation (Lancaster *et al.*, 2000). Since subglottal pressure drops during each echolocation call, it would be expected that two or more pulses close together should have reduced amplitudes since subglottal pressure would not have time to recover in between pulses. Such a scheme seems plausible given that the output energy for two double pulses produced in a row is reduced compared to a single pulse followed by a double pulse [Figs. 6(c) and 7(c)]. The energy reduction is however most marked in the first pulse of the double pulse following a single pulse, and seems to recover amplitude in the second. Such a result implies that it is not subglottal pressure that constrains the amplitude in this pulse but a deliberate partitioning of energy between two pulses.

Compared between runs, a double pulse followed by another double pulse shows reduced intensity, energy duration and bandwidth compared to either single pulses or a double pulse following a single pulse. These consecutive double pulses were rarer than the normal scheme of alternating single and double pulses, but can be seen in Fig. 4(a) where two sets of short interpulse intervals occur at 50 and 70 cm from the target. The reduction in pulse interval, amplitude, and duration is usually indicative of the transition from search to approach phase where the bat has detected the target and begins its approach (Kalko, 1995; Kalko and Schnitzler, 1989). In our experiments, two double pulses may indicate the bat is entering an approach phase which is then aborted.

The bats in these experiments echolocated at around 106 dB peSPL at 10 cm, consistent with recordings in the lab from a variety of species (Surlykke *et al.*, 1993; Waters and Jones 1995). In the field, however, source levels may reach over 120 dB peSPL (Jensen and Miller, 1999; Holderied and von Helversen, 2003; Holderied *et al.*, 2005). In laboratory situations, where bats are not echolocating at full volume, it seems odd that they should still conserve energy by reducing amplitude when the repetition rate rises.

Research has suggested that echolocation during flight requires little energy expenditure above that required for flight alone (Speakman *et al.*, 1989; Speakman and Racey, 1991). The finding that bats may adjust the energy content of pulses, depending on how many pulses are produced per wingbeat cycle, suggests that signal production does incur some degree of energetic expenditure above that required for

flight. *Pipistrellus* and other bat species do sometimes omit echolocation calls during one full wingbeat cycle (Kalko, 1994; Holderied and von Helversen, 2003) implying that there may be some energetic benefit to not echolocating.

If echolocation incurred no extra cost beyond that necessary for flight alone, then one would not expect a reduction in the energy content of signals, even with the emission of more than one signal per wingbeat cycle. Potentially, if the bat has a limited amount of energy to expend per wingbeat cycle and produces two pulses per wingbeat, it must reduce the energy of the pulses in order to stay within its energy budget. Alternatively, the reduction in energy of a double pulse may be the result of a physiological limitation where the bat may not be able to produce two pulses with the same amount of energy as a single pulse in the same breath.

P. pygmaeus exhibits considerable echolocation call plasticity and can adjust their temporal pattern of emission to meet the acoustical demands associated with different tasks. In order to achieve the higher repetition rate associated with echolocation toward a target, the bats increase the production of double pulses per wingbeat cycle. During the emission of double pulses, the energy content per pulse (which presumably reflects the cost of echolocation) is lowered, resulting in two pulses with an energy content 119% of that contained within single pulses produced in consecutive wingbeats. Therefore, an increase in repetition rate during approach flight, by way of increasing double pulse production, probably does not involve a very dramatic, if any, increase in energy expenditure over that required for search flight. *P. pygmaus* may produce an alternation of single and double pulses as part of a preprogrammed sequence, in the same way that Nyctalus noctula alternates different call types (Jones, 1995). However, because the energy content of the first pulse of a double pulse is lowered, the bat may decide whether or not to produce a double pulse in the preceding full pulse, and the production of a double pulse may therefore indicate target detection by the bats from the echo of the preceding pulse.

ACKNOWLEDGMENTS

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).
Bazley, E. N. (**1976**). "Sound absorption in air at frequencies up to 100 kHz," Acoustics Rep. No. AC 74 (National Physics Laboratory, Teddington, UK), pp. 1–43.
Boonman, A., and Jones, G. (**2002**). "Intensity control during target approach in echolocating bats; Stereotypical sensori-motor behaviour in Daubenton's bats, *Myotis daubentonii*," J. Exp. Biol. **205**, 2865–2874.
Britton, A. R. C., Jones, G., Rayner, J. M. V., Boonman, A. M., and Verboom, B. (1997). "Flight performance, echolocation and foraging behaviour in pond bats, *Myotis dasycneme* (Chiroptera: Vespertilionidae)," J. Zool. **241**, 503–522.
Denzinger, A., and Schnitzler, H. U. (**1994**). "Echo SPL influences the ranging performance of the big brown bat, *Eptesicus fuscus*," J. Comp. Physiol., A **175**, 563–571.
Fattu, J. M., and Suthers, R. A. (**1981**). "Subglottic pressure and the control of phonation by the echolocating bat, *Eptesicus fuscus*," J. Comp. Physiol. **143**, 465–475.
Grinnell, A., and Griffin, D. (**1958**). "The sensitivity of echolocating bats," Biol. Bull. **114**, 10–22.
Gustafson, Y., and Schnitzler, H. U. (**1979**). "Echolocation and obstacle avoidance in the hipposiderid bat *Asellia tridens*," J. Comp. Physiol. **131**, 161–167.
Hartley, D. J. (**1992a**). "Stabilization of perceived echo amplitudes in echolocating bats.2. The acoustic behavior of the big brown bat, *Eptesicus fuscus*, when tracking moving prey," J. Acoust. Soc. Am. **91**, 1133–1149.
Hartley, D. J. (**1992b**). "Stabilization of perceived echo amplitudes in echolocating bats. I. Echo detection and automatic gain control in the big brown bat, *Eptesicus fuscus*, and the fishing bat, *Noctilio leporinus*," J. Acoust. Soc. Am. **91**, 1120–1132.
Herbert, H. (**1985**). "Echolocation behavior in the megachiropteran bat, *Rousettus aegyptiacus*," Z. Saugetierkd.-Int. J. Mamm. Biol. **50**, 141–152.
Holderied, M. W., Korine, C., Fenton, M. B., Parsons, S., Robson, S., and Jones, G. (**2005**). "Echolocation call intensity in the aerial hawking bat *Eptesicus bottae* (Vespertilionidae) studied using stereo videogrammetry," J. Exp. Biol. **208**, 1321–1327.
Holderied, M. W., and von Helversen, O. (**2003**). "Echolocation range and wingbeat period match in aerial-hawking bats," Proc. R. Soc. London, Ser. B **270**, 2293–2299.
Jen, P. H. S., and Kamada, T. (**1982**). "Analysis of orientation signals emitted by the Cf-Fm Bat, *Pteronotus p. parnellii* and the Fm Bat, *Eptesicus fuscus* during avoidance of moving and stationary obstacles," J. Comp. Physiol. **148**, 389–398.
Jensen, M. E., and Miller, L. A. (**1999**). "Echolocation signals of the bat *Eptesicus serotinus* recorded using a vertical microphone array: Effect of flight altitude on searching signals," Behav. Ecol. Sociobiol. **47**, 60–69.
Joermann, G., and Schmidt, U. (**1981**). "Echolocation in the common vampire bat, *Desmodus rotundus*.2. Sound emission during flight and correlation with wing beat," Z. Saugetierkd.-Int. J. Mamm. Biol. **46**, 136–146.
Jones, G. (**1995**). "Flight performance, echolocation and foraging in Noctule bats, *Nyctalus noctula*," J. Zool. **237**, 303–312.
Kalko, E. K. V. (**1989**). "The echolocation and hunting behavior of Daubenton's bat, *Myotis daubentoni*," Behav. Ecol. Sociobiol. **24**, 225–238.
Kalko, E. K. V. (**1994**). "Coupling of sound emission and wingbeat in naturally foraging European pipistrelle bats (Microchiroptera, Vespertilionidae)," Folia Zool. **43**, 363–376.
Kalko, E. K. V. (**1995**). "Insect pursuit, prey capture and echolocation in pipistrelle bats (Microchiroptera)," Anim. Behav. **50**, 861–880.
Kalko, E. K. V., and Schnitzler, H. U. (**1989**). "The echolocation and hunting behaviour of Daubenton's bat," Behav. Ecol. Sociobiol. **24**, 225–238.
Kinnear, P. A., and Gray, C. D. (**1999**). *SPSS for Windows Made Simple*, 3rd ed. (Psychology Press, Hove).
Lancaster, W. C., Henson, O. W., and Keating, A. W. (**1995**). "Respiratory muscle-activity in relation to vocalization in flying bats," J. Exp. Biol. **198**, 175–191.
Lancaster, W. C., Keating, A. W., and Henson, O. W. (**1992**). "Ultrasonic vocalizations of flying bats monitored by radiotelemetry," J. Exp. Biol. **173**, 43–58.
Lancaster, W. C., and Speakman, J. R. (**2001**). "Variations in respiratory muscle activity during echolocation when stationary in three species of bat (Microchiroptera: Vespertilionidae)," J. Exp. Biol. **204**, 4185–4197.
Lancaster, W. C., Ward, S., Jones, G., and Speakman, J. R. (**2000**). "Energetics of biosonar vocalization in stationary insectivorous bats," Am. Zool. **40**, 1094–1095.
Moss, C. F., Bohn, K., Gilkenson, H., and Surlykke, A. (**2006**). "Active listening for spatial orientation in a complex auditory scene," PLoS Biol. **4**, 615–626.
Moss, C. F., and Sinha, S. R. (**2003**). "Sonar signal temporal patterning shapes echo-delay tuning in the bat midbrain," 26th Midwinter Research Meeting of the Association for Research into Otolaryngology, Daytona, Beach, FL, 22–27 February, abstract 975.
Neuweiler, G. (**2000**). *The Biology of Bats* (Oxford University Press, Oxford).
Prestwich, K. N. (**1994**). "The energetics of acoustic signaling in anurans and insects," Am. Zool. **34**, 625–643.
Prestwich, K. N., and O″Sullivan, K. (**2005**). "Simultaneous measurement of metabolic and acoustic power and the efficiency of sound production in two mole cricket species (Orthoptera: Gryllotalpidae)," J. Exp. Biol. **208**, 1495–1512.
Simmons, J. A., Moffat, A. J. M., and Masters, W. M. (**1992**). "Sonar gain

control and echo detection thresholds in the echolocating bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **91**, 1150–1163.

Speakman, J. R., Anderson, M. E., and Racey, P. A. (**1989**). "The energy-cost of echolocation in pipistrelle bats (*Pipistrellus pipistrellus*)," J. Comp. Physiol., A **165**, 679–685.

Speakman, J. R., and Racey, P. A. (**1991**). "No cost of echolocation for bats in flight," Nature (London) **350**, 421–423.

Sullivan, W. E. (**1982**). "Possible neural mechanisms of target distance coding in the auditory-system of the echolocating bat *Myotis lucifugus*," J. Neurophysiol. **48**, 1033–1047.

Surlykke, A., Futtrup, V., and Tougaard, J. (**2003**). "Prey-capture success revealed by echolocation signals in pipistrelle bats (*Pipistrellus pygmaeus*)," J. Exp. Biol. **206**, 93–104.

Surlykke, A., Miller, L. A., Mohl, B., Andersen, B. B., Christensendalsgaard, J., and Jorgensen, M. B. (**1993**). "Echolocation in two very small bats from Thailand—*Craseonycteris thonglongyai and Myotis siligorensis*," Behav. Ecol. Sociobiol. **33**, 1–12.

Suthers, R. A. (**1988**). "The production of echolocation signals by bats and birds," in *Animal Sonar: Processes and Performance*, edited by P. E. Nachtigall, and P. W. Moore (Plenum, New York), pp. 23–45.

Suthers, R. A. (**2004**). "Vocal mechanisms in birds and bats: A comparative view," An. Acad. Bras. Cienc. **76**, 247–252.

Suthers, R. A., Thomas, S. P., and Suthers, B. J. (**1972**). "Respiration, wingbeat and ultrasonic pulse emission in an echo-locating Bat," J. Exp. Biol. **56**, 37–48.

Waters, D. A., and Jones, G. (**1995**). "Echolocation call structure and intensity in 5 species of insectivorous bats," J. Exp. Biol. **198**, 475–489.

Wong, J. G., and Waters, D. A. (**2001**). "The synchronisation of signal emission with wingbeat during the approach phase in soprano pipistrelles (*Pipistrellus pygmaeus*)," J. Exp. Biol. **204**, 575–583.

Zar, G. H. (**1999**). *Biostatistical Analysis*, 4th ed. (Prentice Hall, Englewood Cliffs, NJ).

D. A. Waters and J. G. Wong: Energy in echolocation pulses of bats

Biosonar signals impinging on the target during interception by big brown bats, *Eptesicus fuscus*

Prestor A. Saillant, James A. Simmons,[a] and Frederick H. Bouffard
Department of Neuroscience, Brown University, Providence, Rhode Island 02912

David N. Lee
Perception in Action Laboratories, Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, Scotland

Steven P. Dear
Department of Neuroscience, Brown University, Providence, Rhode Island 02912

Big brown bats (*Eptesicus fuscus*) were videotaped in the dark with a night-vision lens and infrared illumination while flying repeatedly along the same straight course to seize a tethered mealworm or a small electret microphone used to record biosonar signals impinging on the target. Bats emitted frequency-modulated sounds with first to third harmonics covering frequencies from 23 to 105 kHz. As the bats neared the target, the first harmonic shifted lower in frequency while the third harmonic strengthened and the fourth harmonic, and sometimes the fifth harmonic, appeared. Incident-sound bandwidth remained broad throughout the maneuver, a feature not seen in field recordings of rapidly moving bats due to propagation losses and uncontrolled directional effects. Sound pressures at the microphone increased by about 20 dB during approach from 2.5 m down to 50 cm and then leveled off, indicating that emitted amplitudes were approximately constant until the terminal stage, when they progressively decreased for the remainder of the maneuver. Interpulse intervals decreased from 80–100 ms down to about 6–7 ms and then stabilized throughout the terminal stage, while durations decreased smoothly from 3–4 ms (limited by adjacent wall) down to 0.5 ms during the terminal stage, which ended with capture. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2714920]

I. INTRODUCTION

Most species of echolocating bats use their sonar to find and intercept flying insects (Fenton, 1995; Griffin, 1958; Neuweiler, 2000; Schnitzler *et al.*, 2003). During aerial interception, the duration and repetition rate of the bat's broadcasts are regulated by the declining delay of echoes from the progressively shortening distance to the target (Griffin *et al.*, 1960; Schnitzler *et al.*, 2003; Simmons *et al.*, 1979). At the terminal stage of the maneuver, a brief, rapid burst of sonar sounds (called a "feeding buzz" from its audible display through an ultrasonic bat detector) guides the final act of interception in which the bat catches its prey by sweeping the tail membrane upward to seize the target (Griffin *et al.*, 1960; Webster and Griffin, 1962). Besides its role for control of pursuit and tracking, the bat's sonar also serves as a perceptual system for perceiving the features of objects (Grinnell, 1995; Moss and Schnitzler, 1995; Neuweiler, 2000; Simmons *et al.*, 1995). For wideband sonar with frequency-modulated (FM) signals, the content of images is based on accurate depiction of the time-of-arrival of each reflected replica of the broadcast sound (Simmons *et al.*, 1995). The critical factor for image quality is the acuity of time resolution for closely spaced reflections, which depends on the

bandwidth of the transmitted signals and the means whereby the bat's auditory system exploits echo bandwidth to register closely spaced delays (Simmons *et al.*, 1996). Perception of delay is affected by changes, even small changes, in the bandwidth and continuity of the frequency band as well as the amplitude of echoes returned by electronic means to the bat's ear's (Masters and Raver, 2000; Moss and Schnitzler, 1989; Simmons *et al.*, 1990, 1998, 2004). Big brown bats can resolve as little as 2 μs in the time separation of echoes from closely spaced reflecting points by registering changes in the frequency of interference notches and peaks in the echo spectrum (Moss and Schnitzler, 1995; Neuweiler, 2000; Simmons *et al.*, 1995). Given the experimental evidence for the big brown bat's high echo-delay acuity in laboratory psychophysical experiments, and the quantitative dependence of this acuity on echo bandwidth, it would be useful to know the extent to which high bandwidth, and, hence, the potential for correspondingly high echo-delay resolution, might be available to bats while they intercept prey.

Answering this question requires knowing the spectral structure of echoes received by bats from targets encountered in nature, which in turn requires characterizing the sounds actually impinging on the targets when the bat makes an attack. In studies of interception, the bat's sounds usually are recorded with microphones located remote from the maneuvering bat and the moving target. The sounds that impinge on the target thus are known only in the most general terms, and

---

[a] Electronic mail: james_simmons@brown.edu

only the most general conclusions are possible about echoes that actually return to the bat's ears. The repetition rate of the bat's emissions can be determined reliably provided the signal levels are even a little above the background noise, and duration can be estimated with some accuracy provided the higher frequencies are not too truncated from the spectrum (a condition rarely met in field recordings), but the full bandwidth and detailed structure of the frequency sweeps in the sounds cannot be known due to directional effects, both at the bat and at the microphone. Atmospheric attenuation especially reduces the high-frequency fidelity of recordings made at microphones located far from the bat (Griffin, 1971; Lawrence and Simmons, 1982). In practice, reliable information is available mainly about the lowest frequencies in the sounds, whereas the highest frequencies also have to be known to delineate the bandwidth and the spectrum.

Methods that bring greater acoustic control over recording conditions include fixing the insect to a specific location relative to the microphone (e.g., Ajrapetjantz and Konstantinov, 1974; Lee *et al.*, 1995; Griffin and Simmons, 1974; Hartley, 1992; Henson *et al.*, 1987; Surlykke and Moss, 2000), employing airborne targets that are deliberately presented in a small region of space to keep the bat within the microphone's receiving beam (Griffin *et al.*, 1965), using multiple microphones so that the bat always has to be pointing at some of them (Ghose and Moss, 2003), or having the bat carry a radio microphone to minimize the effects of movement from one sound to the next (Hiryu *et al.*, 2005). Microphone arrays have been used in the field to examine characteristics of signals emitted at long distances (Jensen and Miller, 1999). In the field or in large laboratory spaces, photographic or video registration of the bat's flight combined with acoustic recordings has been used to reconstruct individual interception maneuvers (Hartley *et al.*, 1989; Holderied *et al.*, 2006; Kalko and Schnitzler, 1998; Schnitzler and Kalko, 1998; see Schnitzler *et al.*, 2003), and several recent studies have applied these methods in the laboratory to reduce the spatial and acoustic variability of the data (Lee *et al.*, 1995; Ghose and Moss, 2003, 2006; Ghose *et al.*, 2006; Surlykke and Moss, 2000). Another way to minimize losses in recording fidelity for flying bats in the wild is to follow the bat's flight with a hand-held thermal-imaging video camera that carries the recording microphone pointing in the same direction as the lens (Eastman and Simmons, 2005; Simmons, 2005; Simmons *et al.*, 2001). This method maximizes the distance at which the bat's sounds can be recorded because the microphone is always pointed toward the bat, but the microphone nevertheless remains an appreciable distance from the bat and insect, and of course it does not register the sound actually impinging on any target. One study addressed this problem by using a stationary bat sitting on a platform while emitting sonar sounds to track a target that was mechanically driven to move nearer (Hartley, 1992). The recording microphone was located on the moving target to register the incident sonar sounds. This method achieved the best possible control over acoustic conditions, but the bat was not actually flying.

In the study reported here, flying bats were trained in repeated trials to capture a mealworm hanging on a fine
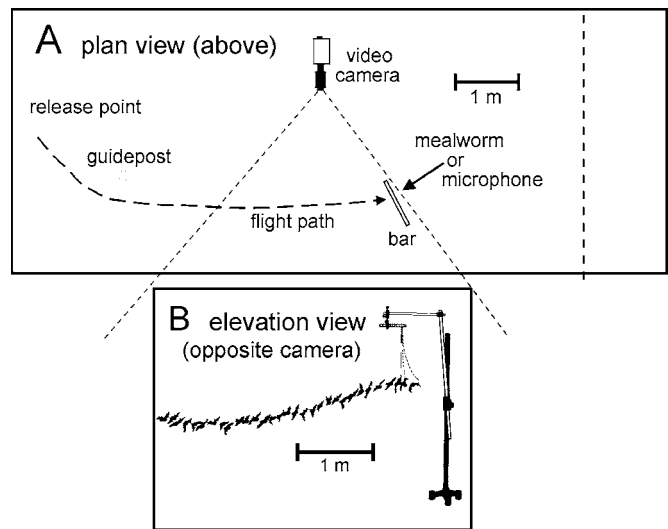
FIG. 1. (A) Diagram of flight room showing position of target (movable) on horizontal bar, with location of video camera and field of view, with guidepost for training bats and approximate flight path. (B) Drawing of video images showing bat's flight and target with support structure. The locations of the bat were traced from the screen of the video monitor by pausing successive video images and outlining the bat's silhouette (17 ms between images). The mealworm is moved to different locations along the horizontal suspension bar to prevent the detailed use of spatial memory during capture. For (B), the 1 m distance scale is only approximate because this tracing has not been corrected for distortions caused by the wide-angle lens.

thread, and then a miniature electret microphone was substituted for the mealworm to record the bat's echolocation sounds during each attack. Video tracking was combined with acoustic analysis to determine the relative sound pressures reaching the microphone during the bat's approach from an initial distance of about 3 m until the microphone was seized by the bat. The goal was to learn how the bat's emitted signals changed in amplitude and frequency structure during aerial interception.

II. METHODS

Big brown bats, *Eptesicus fuscus* (Mammalian order Chiroptera, family Vespertilionidae; see Kurta and Baker, 1990) were trained to repeatedly fly over the same flight-path as they approached and seized a tethered mealworm while the flight was videotaped and the ultrasonic echolocation sounds were recorded (Lee *et al.*, 1995). By replacing the mealworm with a small (expendable) electret condenser microphone, we obtained recordings of the sonar sounds actually impinging on the target during attacks by bats. The bats were collected from attics of houses in Rhode Island and maintained in the laboratory on a diet of mealworms (*Tenebrio* larvae) and water supplemented with vitamin drops (Poly-Vi-Sol). They were housed in a specially equipped colony room kept at 22–25 °C. Air exhaust from the room was through virus-controlling Hepa-filters, and all personnel in contact with the bats were vaccinated against rabies.

A. Flight procedure

Figure 1(a) shows a plan of the flight test room, which was 10.4 m long, 4.2 m wide, and 2.6 m high. The walls, ceiling, and part of the floor were lined with convoluted

polyurethane foam panels (Sonex) to attenuate background echoes. At one end of the room [the far end to the right of Fig. 1(a) as the bat approached the mealworm], nylon mesh netting blocked the bat's access beyond 9 m. The space leading up to the tethered mealworm was slightly less than 6 m [from left to right in Fig. 1(a)], leaving about 3 m past the location of the target for the bat to maneuver with the mealworm in its mouth. The video camera was situated on the bat's left side [Fig. 1(a)] to provide a view of the target and the approaching bat over a span of 2.5–3 m.

Once each bat was trained to fly repeatedly along the established course [Fig. 1(a)], locate the mealworm, and then seize and eat it, successive flight trials were conducted in which the bat was videotaped and its sonar sounds recorded. The elevation view in Fig. 1(b) shows the apparatus for suspending the mealworm and the track from a representative flight by one of the bats traced from the video monitor at 60 fields/s (that is, 60 images/s; see the following), with 17 ms between images. The video monitor displayed 2.5–3 m of the bat's flight path, including a long, straight segment of the bat's approach from the far end of the room as well as the region around the mealworm and its support. The vertical support pole was 1.7 m high, and an attached, slightly inclined movable 1 m bar held a second 75 cm horizontal bar at a height of about 1.7–1.8 m above the floor. Approximately 12 cm below this horizontal bar was a horizontal wooden rail 90 cm long that supported the tethered mealworm. Small lumps of clay held in place with short nails were spaced at 3.5 cm intervals along this bar to provide 36 alternate locations for hanging the nylon filament (0.06 mm diameter) with the mealworm. This arrangement allowed the filament which suspended the mealworm 1.2–1.5 m above the floor to be moved easily from one location to another between flights, and the horizontal rail itself could be oriented in different directions by a swivel on the clamp which attached this rail to the overhead horizontal bar. By this means the exact location of the mealworm was varied from trial-to-trial so that the bat could not affect successful captures just from memory of the location of the pole in the room. The microphone that substituted for the mealworm on measurement trials also was suspended from this bar. It hung from three thin wires which supplied its power and signal connections; these wires were stiff enough that each time the microphone was swung out of the way and then returned, the microphone occupied a different location over a circle about 7–10 cm in diameter from one trial to another.

B. Recording procedure

Each approach flight and capture of the target was videotaped at 30 frames/s (60 fields/s, or 60 images of the bat per second) using a single SVC-09 Xybion video camera equipped with a 1/500 s mechanical shutter. To record at the low light level desired for observing bats while limiting their possible use of vision, the video camera used a Javelin first generation night-vision device (image intensifer) with a wide-angle lens to view the bat. The scene was illuminated for the camera with a 75 W incandescent light in a darkroom fixture fitted with an infrared filter, and visible illumination

was faint. To play back the recordings for analysis, we used a stop-action Gyyr time-lapse recorder, which permitted the video monitor to display each video field to achieve an effective video rate of 60 images/s. The camera was positioned to the bat's left and aimed perpendicular to the bat's flight-path [Fig. 1(a)] to cover the mealworm, most of the mealworm support structure, and about 2–3 m of the bat's approach [Fig. 1(b)]. It was located roughly 2 m from the bat's path and about 1 m above the floor. We trained each bat to fly toward the target along a straight course with minimal lateral movements, making successive repetitions of as nearly the same flight-path as possible given the deliberate trial-to-trial variations in the location of the target.

The bat's sonar broadcasts were recorded with two microphones. One was an ultrasonic condenser microphone (QMC, now Ultra Sound Advice, model SM-2; frequency response ±3 dB from 20 to 110 kHz) placed 1 m past the target and aimed to line up with the bat's flight-path and the aim of its head on the approach. (This microphone was used to obtain clean spectrograms of the bat's sounds for purposes of display, without the low-frequency bias of the Knowles electret microphone.) Then, after several weeks of training (during which video recordings were made for other purposes), on interspersed trials (about a third of the 15–18 trials done on a given day), a small electret microphone (Knowles Model BT-1759; size 8 mm×5 mm×2 mm) was substituted for the tethered mealworm (which was 10–14 mm long and 3–4 mm thick), and sound recordings were made as the bat approached and captured the microphone rather than the mealworm. This microphone is roughly the same size as a mealworm, and subsequent acoustic measurements revealed its target strength to be within 4 dB of the target strength for a mealworm at most orientations (Simmons and Chen, 1989). Its frequency response is ±2 dB from 15 to 20 kHz, then 12 dB/octave rolloff from 20 to 100 kHz. However, over the 20–100 kHz frequency range, the microphone's output noise is declining, too, so practical sensitivity (i.e., signal-to-noise ratio) to external sounds remains approximately flat (±4 dB) from 20 to at least 100 kHz. It has been found to have adequate high-frequency response even for broadband time-of-arrival measurements of the bat's location in flights outdoors (Holderied et al., 2006).

Suspension of the microphone was by a thin, braided 3-lead cable (each wire 0.3 mm in diameter) that supplied the microphone's internal amplifier with power (3.3 V) and retrieved the microphone's output for recording. The electret microphone's front face was oriented by twisting the cable until it pointed approximately toward the bat on its approach flight. This cable was thicker than the nylon filament holding the mealworm, but there was no indication that it affected the flights—the bats intercepted either the microphone or the mealworm with apparent indifference. Even the nylon filament by itself was detectable by the bat because on some flights the mealworm had not yet been placed at the filament's end and the bat attacked the filament's tip. Overall, the bats' catching behavior was very robust. During flights, the bat usually seized the microphone and bit it, making a recognizable sound on the recording before spitting it out and continuing the flight. The bats seemed untroubled by

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Saillant et al.: Bat biosonar sounds at target   3003

being offered the inedible microphone and were willing to make repeated flights with the target alternating between the mealworm and the microphone over the course of each day's tests. They showed no sign of rejecting the microphone after repeated experience, and the sounds recorded by the QMC microphone located 1 m away were similar for both targets.

Signals picked up by the electret microphone were amplified and bandpass-filtered between 10 and 100 kHz (Wavetek/Rockland variable bandpass filter, model 442) before being recorded on three separate channels of a Racal Store-4D tape recorder operating at 76 cm/s. Each channel had a different gain setting to best cover the frequencies and amplitudes of the bat's sounds, which would be expected to increase by roughly 40 dB from the beginning of the bat's approach to the end. The electret microphone's low-pass rolloff did not significantly disrupt the ability to trace the spectrogram of the bat's FM sonar sounds from the tape recordings because the broadcasts of *Eptesicus* have an exceptionally high peak-to-peak amplitude of 100 dB SPL or more, which renders frequencies attenuated up to 20–30 dB by the microphone's frequency response still useful in the recordings. However, making amplitude measurements on the sounds picked up by the electret microphone is more difficult because the microphone's frequency response changes over the 20 to 100 kHz band and because the bat changes the spectrogram of its sounds from the approach to the terminal stage of pursuit. Some frequency regions in the bat's broadcasts thus change in amplitude during the course of the flight because they are shifted from one location to another in the harmonic structure of the broadcasts. These frequencies cannot reliably be used to follow amplitude changes caused by the progressively declining distance from the bat to the microphone. We measured the strength of the bat's sounds in the 25–30 kHz frequency band because this segment of the spectrum is least affected by broadcast directionality (Hartley and Suthers, 1989) and atmospheric attenuation (Griffin, 1971; Lawrence and Simmons, 1982), and because this region of the spectrogram stayed in the first harmonic of the broadcasts to provide a reliable indication of the increase in signal strength at the microphone as the bat flew closer. The signals from the QMC microphone were recorded on the fourth channel of the Racal tape recorder. Additionally, a QMC-mini "bat detector" (which heterodynes the ultrasonic frequencies in the bat's sounds down to lower frequencies) tuned to 30 kHz was used to generate audio-frequency signals which were recorded on the sound track of the videotape to register each sound within the 17 ms duration of its corresponding video field. Synchronization of audio and video signals was achieved by working backwards from the point of capture using the audio track of the videotape as the time reference. The bats would willingly perform 20–30 flights per day if given the opportunity, although potential measurement trials with the electret microphone were limited to 5–7 trials per day.

C. Data analysis

Video images of the bat's flight were analyzed on a video field-by-field basis (60 Hz) in two dimensions corresponding to the length of the flight-path and the bats' height above the floor. A two-dimensional analysis was sufficient for our purposes because the width (bat's left-right) dimension of the flight-path from trial-to-trial was constrained by the straight approach path that the bat had learned to follow. Position information was obtained by pausing each video image on the monitor screen and digitizing the location of the bat's head and the target using a rectangular grid superimposed on the monitor screen. We then transformed video screen coordinates into room coordinates through the use of a videotaped distance calibration grid on the wall that corrected for distortion of images caused by the video camera's wide-angle lens. These data from the bat's flight-path then were merged with corresponding data about when each sonar broadcast occurred to create a data set describing the bat's flight.

For each flight, we played the train of sonar broadcasts recorded by the electret microphone into a Unigon real-time spectrograph to produce a printed spectrogram record showing the timing and FM structure of the sounds. The distinctive sound recorded from the bat's contact with the microphone at the end of the capture maneuver provided convenient registration of the acoustic (Racal) and video recordings, and the timing of individual sounds on the audio track of the video tape was used to corroborate this synchronization. Amplitude envelopes of the same sounds from the Racal tape-recorder were separately plotted after the train of sounds was bandpass filtered at 30 kHz (3 dB bandwidth 4 kHz) rectified, smoothed by a 10 kHz low-pass filter, and digitized using an RC Electronics Type ISC-16 data-acquisition board operating in an IBM PC-AT computer. This same band of frequencies was also used to measure inter-pulse intervals, while the entire broadcast bandwidth was needed for measuring pulse durations. Each sound's duration was measured from the length of the first harmonic (~20–50 kHz) in the printed spectrograms because the frequency response of the microphone posed no problems for its reception. Higher harmonics also are faithfully registered by the Knowles microphone, but their onset times can be affected by the overall 100–105 kHz high-frequency cut-off of the bat's vocal tract. Critical acoustic parameters of the broadcasts picked up by the electret microphone (duration, repetition-rate or interpulse interval, amplitude) were plotted as a function of distance to the target or time time-to-capture as determined from the video recordings. Subsequent analysis of the sounds was done with routines in MATLAB V5.3.

III. RESULTS

Two *Eptesicus* learned to follow the same desired straight flight-path and capture tethered mealworms on repeated trials. They also captured the electret microphone when it was substituted for the mealworm and showed no evident displeasure at biting this inedible object because the bats always continued making subsequent flights to capture mealworms. These bats maintained their willingness to fly and capture suspended mealworms when given flight tests for several months. For detailed analysis of recorded signals, 11 flights were selected for Bat 1 and 14 flights for Bat 2.

FIG. 2. Flash photograph showing a big brown bat flying up to a tethered mealworm. Note the orientation of the bat's head and tail as it prepares to capture the target (photo by S. P. Dear).

These trials were interspersed over about a week of recordings to preclude any limitations associated with a single day's flights. They represent the best examples in which the bats' flight paths kept to a straight approach to avoid directional effects either of the broadcast or of the microphone. Figure 2 illustrates a bat in the act of capturing the mealworm. During trials, the bat flies up to the mealworm on its approach, scoops the worm up with its tail membrane (uropatagium), transfers the worm into its mouth, and begins chewing. Then, the bat continues chewing the worm while it flies around the room or else lands on the wall and finishes the worm at its leisure. The sound of the bat chewing the mealworm always was very audible and served as one indication that a successful capture had occurred. The bat's capture of the microphone could be distinguished by the metallic sound of the bite, which was not always audible to the listener but was especially evident in the recorded signal. Figure 1(b) shows an individual flight that was traced directly from the screen of the video monitor showing silhouettes of the bat in 32 successive video fields at 60 images/s. The pattern of flight shown in Fig. 1(b) and the posture shown in Fig. 2 are typical of the behavior of bats trained for these observations.

A. Flight paths and velocities

Figures 3(a) and 3(c) show flight-paths used by the two *Eptesicus* to capture the tethered electret microphone (11 flights for Bat 1; 14 flights for Bat 2). These tracks correspond to the view shown in Fig. 1(b), but the measurements have been converted from screen coordinates to room coordinates in centimeters relative to the location of the target (at zero on horizontal and vertical axes) to remove the spatial distortion caused by the wide-angle lens and to transform each flight into target-centered units (the actual location of the target varied from trial-to-trial). As the bat approached the stationary microphone, it followed a stereotyped flight-path curving slightly upwards toward the target from below. Figures 3(b) and 3(d) show histograms of the bats' mean horizontal velocities for all of the flights (11 flights of Bat 1 and 14 flights of Bat 2). Overall, Bat 2 flew faster, at about 4.2 m/s, while Bat 1 flew at about 3.5 m/s. For both bats together, mean horizontal velocities ranged from 3.2 to 4.4 m/s on different flights, with enough breadth in the distributions for their ranges to overlap. These average velocities are within the range of flight speeds that have been reported previously for bats (Norberg, 1998; Norberg and Rayner, 1987).

B. Sonar vocalizations

During each flight, the bat emitted a train of FM echolocation sounds which were recorded by the electret microphone that had been substituted for the mealworm. Figure 4(a) compares the spectrum of a bat sound recorded by the Knowles microphone with the spectrum of the same bat sound recorded by the QMC microphone. (These sounds were recorded on different tracks of the Racal recorder with different gains, which were adjusted to keep the strongest sounds just under the limit for distortion of the signals.) Differences between these spectra are due primarily to the frequency response of the Knowles microphone, which rolls off at about 12 dB/octave above 20 kHz, whereas the QMC microphone has essentially a flat response. When this particular sound was emitted, the bat was about 1 m from the target microphone (Knowles) and 2 m from the QMC microphone, which introduced a net additional reduction of 2−4 dB for
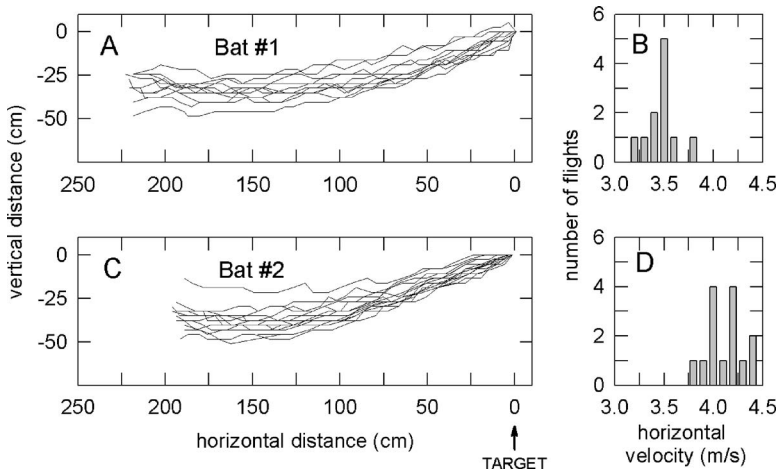
FIG. 3. (A), (C) Flight trajectories and (B), (D) mean horizontal velocities for two bats during interception maneuvers with the microphone as the target (11 trials for Bat 1 and 14 trials for Bat 2). Horizontal and vertical axes show distances relative to the target (arrow) after conversion from screen units to centimeters in the room.
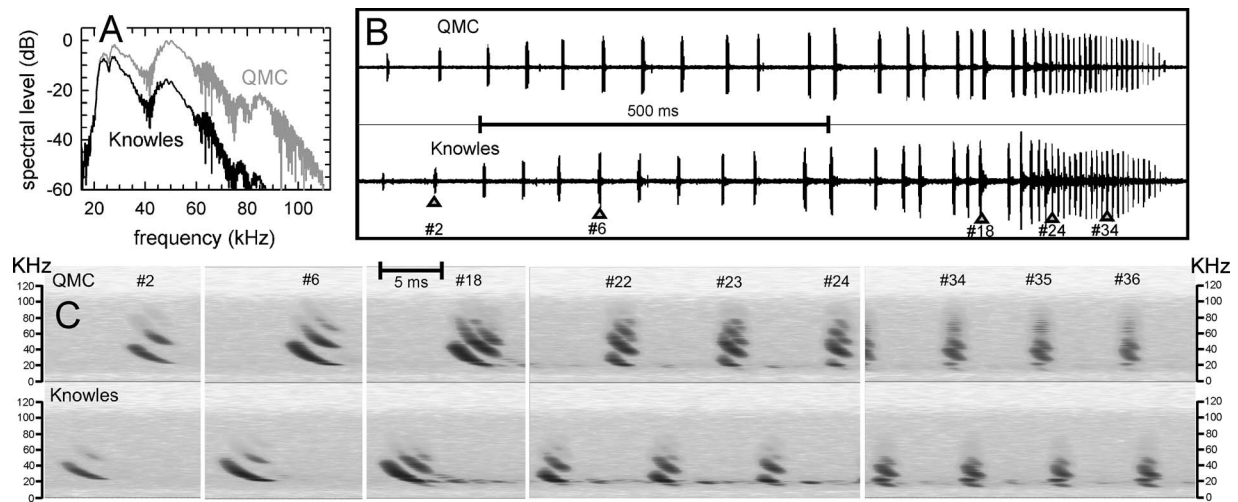
FIG. 4. (A) Spectra for a typical big brown bat echolocation sound recorded with the QMC and Knowles microphones. (B) Time-series plots of echolocation sounds recorded with the QMC and Knowles microphones during a single interception flight by Bat 1. (C) Spectrograms of nine sounds selected from the series in (B).

50–100 kHz at the QMC microphone due to atmospheric absorption (Griffin, 1971; Lawrence and Simmons, 1982). Figure 4(b) shows the time-series wave forms from one complete interception flight by Bat 1 recorded by the QMC and Knowles microphones. The bat's signals arrived 6 ms later at the QMC microphone, which is 1 m further away. Aside from differences in the background noise and the amplitude profile from sound-to-sound [plus spectral differences as in Fig. 4(a)], the signal trains appear alike. Figure 4(c) compares spectrograms for nine of the signals recorded at the two microphones (sounds 2, 6, 18, 22, 23, 24, 34, 35, 36). In these spectrograms, the first and second harmonic sweeps are equally prominent for both QMC and Knowles recordings, but the third and fourth harmonics show up only in the QMC spectrograms on account of the low-pass rolloff of the Knowles microphone [see Fig. 4(a)]. Note that the first harmonic, which was used to estimate signal duration, is present in its entirety in the Knowles recordings. If a horizontal line is drawn across the top of either the QMC or the Knowles series of spectrograms from No. 2 to at least No. 24, the upper frequency limit is seen to be about the same across the series.

Figure 5 shows spectrograms of 42 successively numbered sounds picked up at the QMC microphone from Bat 1 during one flight to capture a mealworm. The individual sounds were taken from different tracks of the recordings, and most were about 75%–90% of the maximum amplitude for causing significant distortion. However, in this series, sounds 18 and 20–23 did exceed the limit for distortion; they are included in the spectrograms to illustrate how the appearance of spurious frequencies above 100 kHz serves as a sign that distortion did occur. All the other sounds were below the distortion limit. The broadcast sounds consisted of three to five descending harmonic sweeps that collectively covered frequencies from a minimum of 15–20 kHz to a maximum of 100–105 kHz. The duration of the sounds declined steadily from about 4 ms to as little as 0.7 ms over the course of the flight. The FM structure of the sounds changed gradually during the maneuver from three distinct harmonic

sweeps in the approach stage (Nos. 1–14) to four harmonic sweeps early in the terminal stage (Nos. 15–25), and to five harmonic sweeps near the end of the terminal stage (Nos. 26–40). These changes are achieved by a progressive slide of the first harmonic to lower frequencies (it sweeps from 55 down to about 25 kHz in No. 4, from 51 down to 21 kHz in No. 19, from 34 down to 18 kHz in No. 27, and from 20 down to 15 kHz in No. 35) combined with corresponding downward shifts of the second and third harmonics. The fourth and fifth harmonics make an appearance to fill the frequencies left empty when the first to third harmonics shifted to lower frequencies. The highest frequency in the sounds thus remained at 100–105 kHz throughout the flight, while the lowest frequency actually decreased somewhat. Otherwise the sounds in Fig. 5 are typical of sonar broadcasts used by *Eptesicus* during pursuit of flying prey (Griffin, 1958; Surlykke and Moss, 2000) or interception of a suspended mealworm (Ghose and Moss, 2003, 2006). They also are similar to sonar broadcasts emitted by stationary bats tracking a microphone that moves toward the bat on a mechanical slide (Hartley, 1992). Furthermore, they are similar to sounds broadcast by *Eptesicus* in numerous behavioral experiments employing target or echo discrimination techniques (Moss and Schnitzler, 1995; Simmons *et al.*, 1995).

Figure 6 shows the duration, interpulse interval, and relative sound pressure at 30 kHz of sonar broadcasts recorded by the electret (target) microphone during all 11 flights by Bat 1 [Figs. 6(a)–6(c)] and all 14 flights of Bat 2 [Figs. 6(d)–6(f)]. These three acoustic parameters are plotted against target range (radial distance to target). As each bat approached the target, it transmitted sounds at a higher repetition rate (shorter interpulse interval; Figs. 6(b) and 6(e)), thus giving a progressively greater density of data-points per-unit-time going from 250 cm on the left to 0 cm. This is not surprising; the increased repetition rate of broadcasts is the most thoroughly documented effect in sonar-guided aerial interceptions (Griffin, 1958; Schnitzler *et al.*, 2003; Simmons *et al.*, 1979). Both bats shifted somewhat raggedly to a faster rate at a distance of 200–225 cm (vertical dashed lines
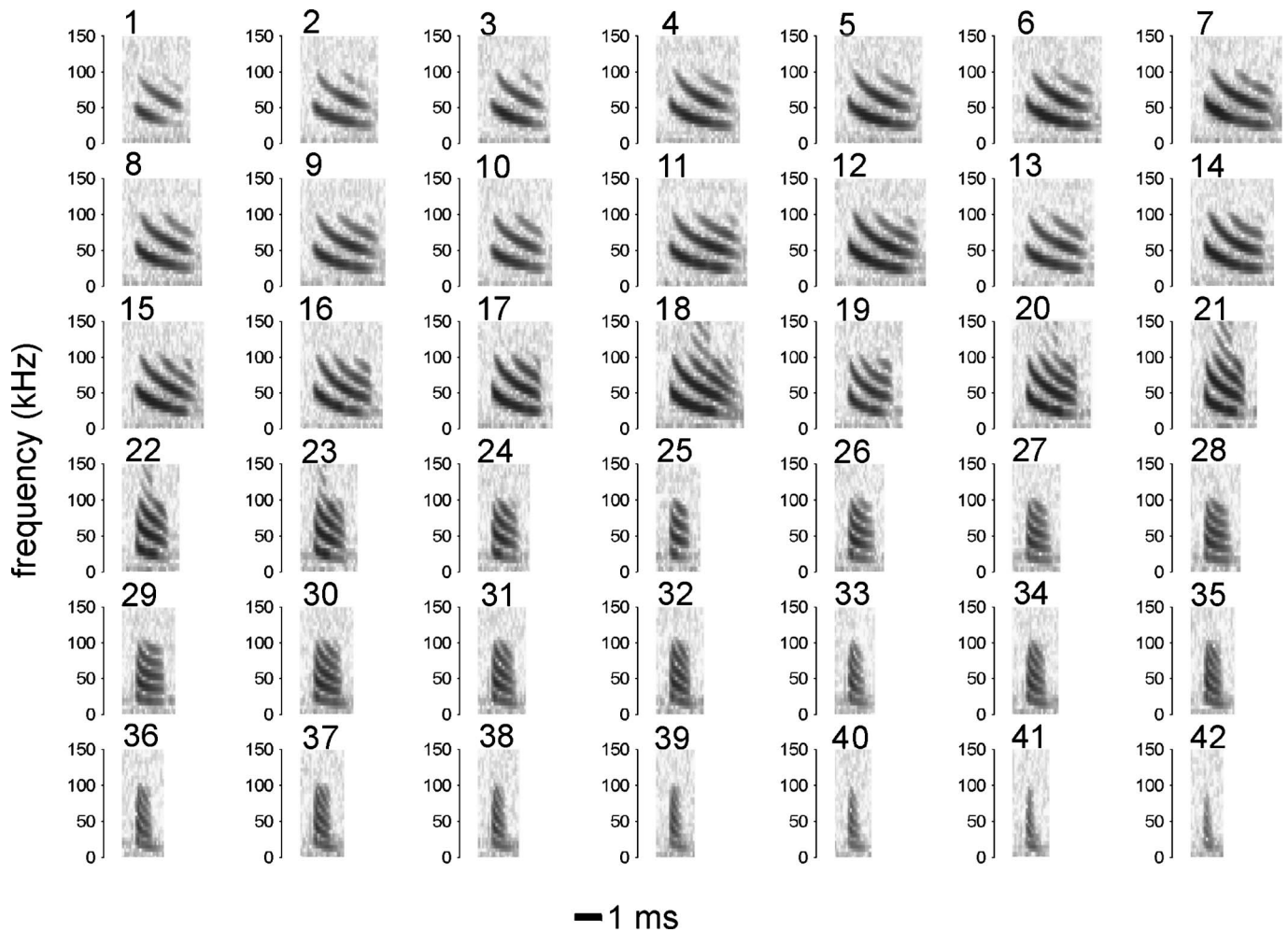
FIG. 5. Spectrograms of 42 sequentially numbered echolocation sounds recorded with the QMC microphone during one approach by Bat 1. Although individual harmonics shift to lower frequencies as the flight progresses, overall, the bandwidth of the sounds remains wide.

on the left in the graphs). The interval between sounds was 90–100 ms when the bat entered the field-of-view of the video camera, at a distance of roughly 3 m from the target (see Fig. 1), and it declined to about 6–7 ms by the time of capture, leveling off at this value during the final 50–75 cm of flight. In Figs. 6(a) and 6(d) both bats kept the duration of their sounds at about 3–3.5 ms (full range of 2–4 ms) until they approached to within 75–100 cm of the target, at which point they progressively decreased the duration to about 0.5 ms at the moment of capture (vertical dashed lines on the right of the graphs). For both bats, the interpulse intervals also became very short and less variable after the target's range has declined to about 50–75 cm. This marks the transition into the terminal stage, where the duration of emissions begins its pronounced decrease from about 2.5 to as little as 0.5 ms (approximately at the location of the right-hand vertical dashed lines).

In Figs. 6(c) and 6(f), the amplitude of the sounds impinging on the target microphone increased as the bats approached from a distance of 2–3 m and then leveled off after about 50–75 cm. These changes are better illustrated in Fig. 7, which shows logarithmic plots to emphasize proportional relations between distance and relative amplitude (dB) values. For Bat 1, amplitudes changed from −15 to −20 dB (relative to 0 dB for the strongest sound recorded in all the flights for each bat) at an initial distance of about 2.5–3.0 m to −3 to −7 dB at distances of 50–75 cm. For Bat 2, the amplitudes changed from about −20 to −27 dB at an initial distance of 2.5–3.0 m to −2 to −10 dB at distances of 50–75 cm. For both bats, at distances in the range from about 60 cm down to 40 cm, the amplitude recorded at the target remained roughly constant (between −2 and −7 dB for Bat 1; between −2 and −10 dB for Bat 2). Throughout the rest of the terminal stage, from 40 down to 10 cm, the amplitude of the incident sounds remained constant for Bat 1 while it declined by about 5 dB for Bat 2. This latter segment occurred while the bats were changing flight posture and head aim from approaching the target head-on to reorienting posture and head so that the tail membrane rotated upward toward the front to capture the target microphone. Figure 2 shows a bat performing this rotation. It is difficult to interpret the constant (for Bat 1) or slightly declining (for Bat 2) amplitudes at distances of 10–30 cm because the bat may be shifting its sonar beam upward in relation to the target.

IV. DISCUSSION

Measurements of big brown bat sonar sounds incident on the target were made with a procedure that transferred the

J. Acoust. Soc. Am., Vol. 121, No. 5, May 2007

Saillant *et al.*: Bat biosonar sounds at target   3007
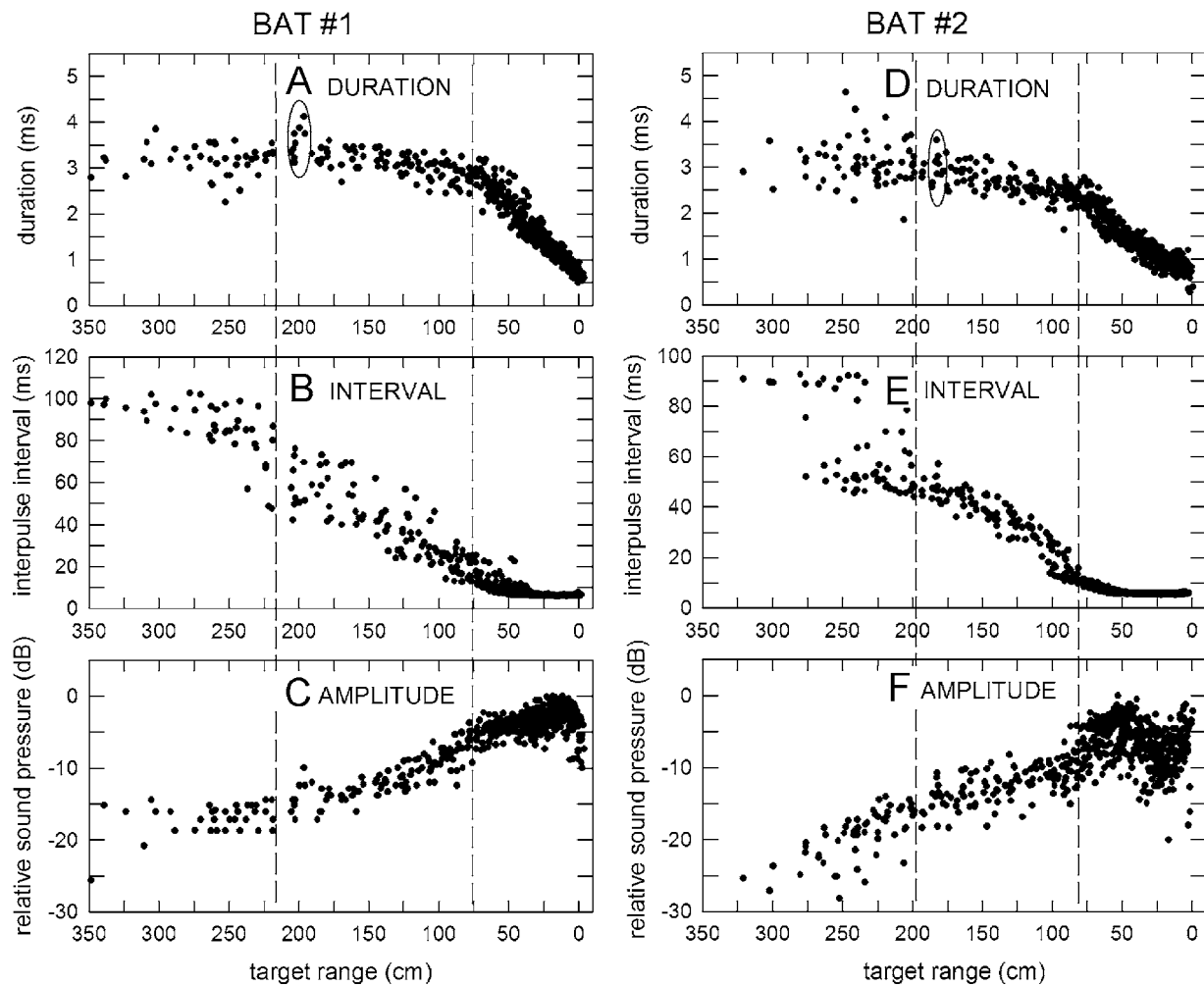
FIG. 6. Plots of (A) duration, (B) interpulse interval, and (C) relative amplitude (sound pressure) for echolocation sounds recorded from the target microphone in all 11 flights of Bat 1. Plots of (D) duration, (E) interpulse interval, and (F) relative amplitude for sounds recorded from the target microphone in all 14 flights of Bat 2.

essential elements of the Hartley (1992) study that used stationary bats and a moving target into a new procedure that used flying bats and a stationary target (Fig. 1). The results reported here cover several aspects of the bat's flight movements (Fig. 3) and vocalizations (Figs. 4–7) in relation to the target. Most parameters of the trains of echolocation sounds recorded here (e.g., pulse interval, pulse duration, pulse amplitude) are similar to values reported previously for big brown bats under less well-controlled acoustic conditions in the field or the laboratory (Griffin, 1958; Hartley, 1992; Simmons *et al.*, 1979; Surlykke and Moss, 2000), and there are no obvious changes necessary in the conclusions drawn from those earlier studies.

However, the procedure used here focused on a question that has not been considered in the context of the previous studies—How does the amplitude and bandwidth of the bat's *incident* sounds change during aerial interception maneuvers? The present method achieved sufficient control over acoustic conditions to determine two new facts: First, the bandwidth of signals emitted during the approach and terminal stages remains approximately the same throughout the maneuver (Fig. 5). Previous studies have described the big brown bat's sounds as having first, second, and sometimes
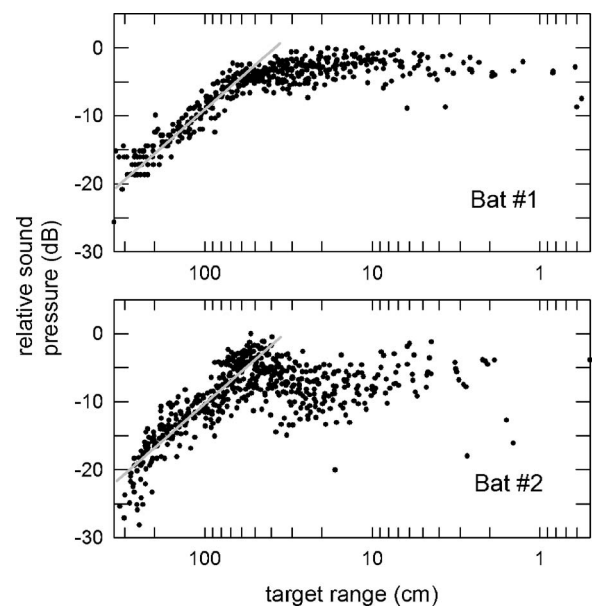
FIG. 7. Logarithmic plots for the relative sound pressures of sounds recorded at the microphone for Bats 1 and 2 [data from Figs. 6(c) and 6(f)]. The gray line in each graph shows the slope of a proportional relation between increasing amplitude and decreasing distance (20 dB increase in sound pressure for factor of 10 reduction in distance).

even third harmonic sweeps during the approach and terminal stages of interception, while most recordings also show the sounds becoming progressively weaker during the terminal stage. It is known that the frequencies in the first and second harmonics shift to lower values during the terminal stage, and field recordings give the impression that higher (third, fourth, and fifth) harmonics progressively drop out of the sounds as the terminal stage proceeds to the moment of capture. Some of this degradation may be due to the differential loss of high frequencies due to atmospheric absorption. Furthermore, the terminal stage often is associated with the most radical maneuvering by the bat to achieve interception of the target, and the emitted sounds would naturally only rarely be aimed directly at the microphone throughout the "feeding buzz." However, recordings made with stationary bats show that the shift of harmonics to lower frequencies in the terminal stage is accompanied by insertion of higher harmonics to maintain the same overall bandwidth (Hartley, 1992). The recordings made here confirm this observation in flying bats. The higher harmonics (third, fourth, fifth)—which would not be well represented in recordings made with microphones located off the main acoustic axis of the bat's broadcasts (see Hartley and Suthers, 1989), or not pointed directly at the bat—still are emitted by the bat, right up to the act of capture so that the upper cutoff frequency is kept fixed at 100–105 kHz (Fig. 5).

The second finding concerns the profile of amplitudes for the incident sounds. When the bat is trained to make its approach so that the target *is* the microphone, so that sound-pressure losses due to long distances and off-axis effects are minimized, the relative amplitudes of the incident biosonar sounds increased progressively during the approach stage and then remain more nearly constant during the terminal stage [Figs. 6(c) and 6(f), Fig. 7]. These amplitude measurements are for the incident sounds arriving at the microphone; what about the amplitudes of the sounds actually emitted by the bat? If the sounds are constant at the bat, moving nearer the microphone by a factor of 10 should lead to a proportional increase in sound pressure by a factor of 10, or 20 dB. The logarithmic plots in Fig. 7 illustrate this relation. The solid gray line in each plot shows the slope for proportional increases in sound pressure in relation to declining distance (20 dB per factor of 10 reduction in distance). From the similarity in the slope of the lines and the distribution of points in Fig. 7, to a first approximation the amplitudes actually emitted by the bat are relatively constant while the bat approaches from 250–300 down to 50–75 cm. In the terminal stage, the recorded amplitudes remain constant or decline slightly, implying that the bats decrease the amplitude of their broadcasts roughly in proportion to the remaining declining distance during the terminal stage. That is, because the bat is moving nearer to the microphone, the roughly constant recorded sound pressure implies a decreasing source sound pressure. In Fig. 7, the only obvious difference between the two bats is the higher scatter of points for Bat 2 compared to Bat 1 and the decline in amplitudes for Bat 2 relative to Bat 1 during the final 30–40 cm of the terminal stage. As Fig. 2 illustrates, the flight behavior of the bats changes during this interval, so it is not possible to know whether the small decrease for Bat 2 reflects real, but small, changes in the emitted sounds or changes in the sound field due to head and posture shifts.

The observation that the bandwidth of the signals remains about the same throughout the approach and terminal stages has an important implication about the bat's perceptions during interception (Hartley, 1992). It seems reasonable to conclude that the bat is trying to perceive with its best accuracy and resolution one or more of the target's features whose perceptual acuity depends on having a broad echo bandwidth, even after the bat has made its decision to proceed with an attempt to capture the target (presumably at the transition from the approach into the terminal stages). These perceived features potentially include distance, shape, horizontal position, and vertical position. Certainly, active tracking of target direction has to be maintained throughout the approach and terminal stages so that interception can take place in spite of changes in the insect's position relative to the bat (Ghose and Moss, 2006; Ghose, *et al.*, 2006). It also is possible that the bat retains its ability to perceive direction, as distinct from tracking direction, and even to discriminate target features such as shape, past the end of the approach. Video recordings of big brown bats in the field show that they are capable of quickly changing their mode of echolocation from the terminal stage back to the approach stage in cases where the insect has been dropped or has taken evasive action, so that the target is reacquired and reintercepted even after the bat has begun the terminal stage (Simmons *et al.*, 2001; Simmons, 2005).

ACKNOWLEDGMENTS

Ajrapetjantz, E. Sh., and Konstantinov, A. I. (**1974**). "Echolocation in Nature," Rep. No. 63328-1,-2, Joint Publications Research Service, Arlington, VA.

Eastman, K. A., and Simmons, J. A. (**2005**). "A method of flight path and chirp pattern reconstruction for multiple flying bats," ARLO **6**, 257–262.

Fenton, M. B. (**1995**). "Natural history and biosonar signals," in *Hearing by Bats*, edited by A. N. Popper, and R. R. Fay (Springer, New York), pp. 37–86.

Ghose, K., and Moss, C. F. (**2003**). "The sonar beam pattern of a flying bat as it tracks tethered insects," J. Acoust. Soc. Am. **114**, 1120–1131.

Ghose, K., and Moss, C. F. (**2006**). "Steering by hearing: A bat's acoustic gaze is linked to its flight motor output by a delayed, adaptive filter law," J. Neurosci. **26**, 1704–1710.

Ghose, K., Horiuchi, T. K., Krishnaprasad, P. S., and Moss, C. F. (**2006**). "Echolocating bats use a nearly time-optimal strategy to intercept prey," PLoS Biol. **4**, 865–873.

Griffin, D. R. (**1958**). *Listening in the Dark* (Yale University Press, New Haven CT; reprinted by Cornell University Press, Ithaca, NY, **1986**).

Griffin, D. R. (**1971**). "The importance of atmospheric attenuation for the echolocation of bats (Chiroptera)," Anim. Behav. **19**, 55–61.

Griffin, D. R., Friend, J. H., and Webster, F. A. (**1965**). "Target discrimination by the echolocation of bats," J. Exp. Zool. **158**, 155–168.

Griffin, D. R., and Simmons, J. A. (**1974**). "Echolocation of insects by horseshoe bats," Nature (London) **250**, 731–732.

Griffin, D. R., Webster, F. A., and Michael, C. (**1960**). "The echolocation of

flying insects by bats," Anim. Behav. **8**, 141–154.

Grinnell, A. D. (**1995**). "Hearing in bats: An overview," in *Hearing by Bats*, edited by A. N. Popper, and R. R. Fay (Springer, New York), pp. 1–36.

Hartley, D. J. (**1992**). "Stabilization of perceived echo amplitudes in echolocating bats. II. The acoustic behavior of the big brown bat; *Eptesicus fuscus*, when tracking moving prey," J. Acoust. Soc. Am. **91**, 1133–1149.

Hartley, D. J., Campbell, K. A., and Suthers, R. A. (**1989**). "The acoustic behavior of the fish-catching bat, *Noctilio leporinus*, during prey capture," J. Acoust. Soc. Am. **86**, 8–27.

Hartley, D. J., and Suthers, R. A. (**1989**). "The sound emission pattern of the echolocating bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **85**, 1348–1351.

Henson, O. W., Jr., Bishop, A. L., Keating, A. W., Kobler, J. B., Henson, M. M., Wilson, B. S., and Hansen, R. C. (**1987**). "Biosonar imaging of insects by *Pteronotus p. parnelli*, the Mustached Bat," Natl. Geog. Res. **3**, 82–101.

Hiryu, S., Katsura, K., Lin, L.-K., Riquimaroux, H., and Watanabe, Y. (**2005**). "Doppler-shift compensation in the Taiwanese leaf-nosed bat (*Hipposideros terasensis*) recorded with a telemetry microphone system during flight," J. Acoust. Soc. Am. **118**, 3927–3933.

Holderied, M. W., Jones, G., and von Helverson, O. (**2006**). "Flight and echolocation behavior of whiskered bats commuting along a hedgerow: Range-dependent sonar signal design, Doppler tolerance and evidence for 'acoustic focusing'," J. Exp. Biol. **209**, 1816–1826.

Jensen, M. E., and Miller, L. (**1999**). "Echolocation signals of the bat *Eptesicus serotinus* recorded using a vertical microphone array: Effect of flight altitude on searching signals," Behav. Ecol. Sociobiol. **47**, 60–69.

Kalko, E. K. V., and Schnitzler, H.-U. (**1998**). "How echolocating bats approach and acquire food," in *Bat Biology and Conservation*, edited by T. H. Kunz and P. A. Racey (Smithsonian Institution Press, Washington D.C.), pp. 197–204.

Kurta, A., and Baker, R. H. (**1990**). "*Eptesicus fuscus*," Mammalian Species **356**, 1–10.

Lawrence, B. D., and Simmons, J. A. (**1982**). "Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats," J. Acoust. Soc. Am. **71**, 585–590.

Lee, D. N., Simmons, J. A., Saillant, P. A., and Bouffard, F. (**1995**). "Steering by echolocation: A paradigm of ecological acoustics," J. Comp. Physiol. [A] **176**, 347–354.

Masters, W. M., and Raver, K. S. (**2000**). "Range discrimination by big brown bats (*Eptesicus fuscus*) using altered model echoes: Implications for signal processing," J. Acoust. Soc. Am. **107**, 625–637.

Moss, C. F., and Schnitzler, H.-U. (**1989**). "Accuracy of target ranging in echolocating bats: Acoustic information processing," J. Comp. Physiol. [A] **165**, 383–393.

Moss, C. F., and Schnitzler, H.-U. (**1995**). "Behavioral studies of auditory information processing," in *Hearing by Bats*, edited by A. N. Popper, and R. R. Fay (Springer, New York), pp. 87–145.

Neuweiler, G. (**2000**). *The Biology of Bats* (Oxford University Press, New York).

Norberg, U. M. (**1998**). "Morphological adaptations for flight in bats," in *Bat Biology and Conservation*, edited by T. H. Kunz and P. A. Racey (Smithsonian Institution Press, Washington, D.C.), pp. 93–108.

Norberg, U. M., and Rayner, J. M. V. (**1987**). "Ecological morphology and flight in bats (Mammalia; Chiroptera): Wing adaptations, flight performance, foraging strategy, and echolocation," Philos. Trans. R. Soc. London, Ser. B **316**, 335–427.

Schnitzler, H.-U., and Kalko, E. K. V. (**1998**). "How echolocating bats search and find food," in *Bat Biology and Conservation*, edited by T. H. Kunz and P. A. Racey (Smithsonian Institution Press, Washington, D.C.), pp. 183–196.

Schnitzler, H.-U., Moss, C. F., and Denzinger, A. (**2003**). "From spatial orientation to food acquisition in echolocating bats," Trends Ecol. Evol. **18**, 386–394.

Simmons, J. A. (**2005**). "Big brown bats and June beetles: Multiple pursuit strategies in a seasonal acoustic predator-prey system," ARLO **6**, 238–242.

Simmons, J. A., and Chen, L. (**1989**). "Acoustic basis for target discrimination by FM echolocating bats," J. Acoust. Soc. Am. **86**, 1333–1350.

Simmons, J. A., Fenton, M. B., and O'Farrell, M. J. (**1979**). "Echolocation and pursuit of prey by bats," Science **203**, 16–21.

Simmons, J. A., Ferragamo, M. J., and Moss, C. F. (**1998**). "Echo-delay resolution in sonar images of the big brown bat, *Eptesicus fuscus*," Proc. Natl. Acad. Sci. U.S.A. **95**, 12647–12652.

Simmons, J. A., Ferragamo, M. J., Moss, C. F., Stevenson, S. B., and Altes, R. A. (**1990**). "Discrimination of jittered sonar echoes by the echolocating bat, *Eptesicus fuscus*: The shape of target images in echolocation," J. Comp. Physiol. [A] **167**, 589–616.

Simmons, J. A., Ferragamo, M. J., Saillant, P. A., Haresign, T., Wotton, J. M., Dear, S. P., and Lee, D. N. (**1995**). "Auditory dimensions of acoustic images in echolocation," in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 146–190.

Simmons, J. A., Neretti, N., Intrator, N., Altes, R. A., Ferragamo, M. J., and Sanderson, M. I. (**2004**). "Delay accuracy in bat sonar is related to the reciprocal of normalized echo bandwidth, or *Q*," Proc. Natl. Acad. Sci. U.S.A. **101**, 3638–3643.

Simmons, J. A., O'Farrell, M. J., and Horowitz, S. S. (**2001**). "Versatility of biosonar in the big brown bat, *Eptesicus fuscus*," ARLO **2**, 43–48.

Simmons, J. A., Saillant, P. A., Ferragamo, M. J., Haresign, T., Dear, S. P., Fritz, J. B., and McMullen, T. A. (**1996**). "Auditory computations for acoustic imaging in bat sonar," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay, Springer Handbook on Auditory Research (Springer, New York), pp. 401–468.

Surlykke, A., and Moss, C. F. (**2000**). "Echolocation behavior of big brown bats, *Eptesicus fuscus*, in the field and the laboratory," J. Acoust. Soc. Am. **108**, 2419–2429.

Webster, F. A., and Griffin, D. R. (**1962**). "The role of the flight membrane in insect capture by bats," Anim. Behav. **10**, 332–340.

Noise level correlates with manatee use of foraging habitats

Jennifer L. Miksis-Olds[a] and Percy L. Donaghay
Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882

James H. Miller
Department of Ocean Engineering & Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882

Peter L. Tyack
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Jeffrey A. Nystuen
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Box 355640, Seattle, Washington 98105

The introduction of anthropogenic sound to coastal waters is a negative side effect of population growth. As noise from boats, marine construction, and coastal dredging increases, environmental and behavioral monitoring is needed to directly assess the effect these phenomena have on marine animals. Acoustic recordings, providing information on ambient noise levels and transient noise sources, were made in two manatee habitats: grassbeds and dredged habitats. Recordings were made over two 6-month periods from April to September in 2003 and 2004. Noise levels were calculated in one-third octave bands at nine center frequencies ranging from 250 Hz to 64 kHz. Manatee habitat usage, as a function of noise level, was examined during four time periods: morning, noon, afternoon, and night. Analysis of sightings data in a variety of grassbeds of equal species composition and density indicate that manatees select grassbeds with lower ambient noise for frequencies below 1 kHz. Additionally, grassbed usage was negatively correlated with concentrated boat presence in the morning hours; no correlation was observed during noon and afternoon hours. This suggests that morning boat presence and its associated noise may affect the use of foraging habitat on a daily time scale. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2713555]

I. INTRODUCTION

Florida manatees live in shallow coastal habitats that typically range from $\frac{1}{2}$ m to approximately 10 m in depth. The deepest portions of the habitat are utilized most often during travel from one site to another and occur in dredged channels such as the Intracoastal Waterway, open bay, and open ocean areas (Koelsch, 1997). When manatees are not traveling, they spend the majority of their time in the shallower waters of seagrass beds, sand bars, and secluded dredged habitats (Marine Mammal Commission, 1986). To better understand how noise may affect habitat usage in manatees, it is necessary to quantify ambient noise levels in their shallow water habitats.

Manatees are thought to rely upon sound for long range communication (Sousa-Lima *et al.*, 2002). Because sound has the potential to travel long distances in water, it provides a reliable way for manatees to communicate beyond visual range in murky coastal waters. Other types of environmental sound signals, such as the breaking of surface waves on seawalls and land, can aid in navigation, whereas sound produced from watercraft can warn of an approaching vessel.

The issue of how sound travels in the shallow waters of manatee habitats has recently become a topic of heightened interest because of the risk of collision between manatees and boats. The question of whether or not manatees can hear boats approaching in enough time to swim to safety is a matter of active debate. Understanding more about the noise levels to which manatees are exposed will not only contribute to answering the question of whether manatees can hear boats approaching, but it will also provide information on how the natural communication system of manatees may be impacted by rising levels of background noise and/or transient noise sources in their habitats.

An acoustic signal can be only detected when the sound intensity level of the signal exceeds the auditory threshold of the individual receiving the sound and is greater than the level of ambient noise over the bandwidth of hearing at similar frequencies. The lower the background noise, the farther a sound signal will travel before its level diminishes below the background noise level (Richardson *et al.*, 1995). The effective range of communication, or range within which the information contained in signal is successfully transmitted to a receiver, depends on the ambient noise levels, acoustic propagation loss characteristics, and frequency and amplitude of the signals being transmitted to the receiver by the producer. Higher frequencies typically have a higher attenuation than lower frequencies, but in very shallow grassbed

---

[a]Current affiliation: School for Marine and Science Technology, University of Massachusetts Dartmouth, 838 S. Rodney French Blvd., New Bedford, MA 02744. Electronic mail: jmiksis@umassd.edu

habitats the most efficient sound propagation occurs in the range of 2–20 kHz (Miksis-Olds and Miller, 2006). This occurs because low frequency sound wavelengths are larger than the water depth, resulting in quick absorption by the sediments. The Lloyd mirror effect also contributes to a greater attenuation in the low frequency component of signals (Etter, 1996). Environmental parameters such as water depth, salinity, temperature, bottom type, and wind speed also affect sound absorption and attenuation. Consequently, sound transmission differs for varying wavelengths in different manatee habitats, and different habitat types therefore affect the range at which manatees can detect either conspecific vocalizations or approaching vessels. It is entirely possible that both environmental noise and transmission loss are so great in some areas that manatees cannot detect boats until they are only a few tens of meters away. Manatees have been shown to utilize grassbeds with higher transmission loss characteristics more often than grassbed sites with lower levels of transmission loss (Miksis-Olds and Miller, 2006). High transmission loss compounded by high levels of ambient noise in grassbeds greatly diminishes the likelihood of detecting an approaching watercraft while it is still in the distance. Measurements of ambient noise levels associated with this study provide details to address whether or not this is a common circumstance.

The probability of detecting signals of interest, such as conspecific vocalizations, can also be reduced by masking from other acoustic signals in the environment. Masking, or the obscuring of sounds of interest by interfering sounds at similar frequencies, may hamper an animal's ability to detect a sound signal even when that signal is above the absolute hearing threshold (Richardson et al., 1995). The potential masking by high levels of human generated noise, as well as the cacophony of sounds manatees naturally encounter, provides a constant obstacle to effective vocal communication and the perception of other biologically significant signals. The level of masking is mainly due to noise at frequencies near the signal frequency of interest. Noise at frequencies outside this masking band has little influence on the detection of a signal unless the noise level is very high (Kryter, 1985; Richardson et al., 1995). From the viewpoint of a manatee, any signal of interest must overcome not only absorption and attenuation, but also masking by background noise and a myriad of broadband and narrow-band sounds. Ambient noise is environmental background noise that is generally unwanted sound that clutters and masks other sounds of interest (Richardson et al., 1995). Ambient noise has components that can be continuous or transient, broadband or tonal, directional or omnidirectional. Some of the major sound components in manatee habitats include high levels of noise due to wind and waves (broadband), watercraft (broadband), snapping shrimp (broadband), and fish and marine mammal vocalizations (broadband and narrow band).

With a firm grasp on both environmental noise and transmission loss in a specific habitat at a particular time, it becomes possible to begin to address questions of the range at which a signal can be detected. As a rule of thumb, a signal must be louder than the background noise level at similar frequencies in order for it to be detected. Conse-

quently, a signal with a high source level and low amount of transmission loss may or may not be detected a mere 100 m away depending on how noisy the environment is. The ambient noise levels in coastal waters, bays, and harbors are subject to wider variations than the deep-water ambient noise, and coastal marine mammals must cope with sources of noise that are highly variable in time, frequency, and space (Urick, 1983). The motivation behind this study was to determine if manatees perceive and react to changes in ambient noise levels. Manatee habitat usage was examined in relation to ambient noise levels.

II. METHODS

Noise recordings were made in two habitat types: seagrass beds and dredged habitats. Sampling sites consisted of 13 grassbeds (sites A, B, D, F, G, H, I, N, O, S, T, V, and X) and 11 dredged habitats (sites C, E, J, K, L, M, P, Q, R, U, W) in Sarasota Bay, FL. The letter designations for each site were assigned based on geographic location in Sarasota Bay starting near Mote Marine Laboratory and progressing in a counter-clockwise direction around the bay (Fig. 1). The site identifications in this study are consistent with previously published work at the same sites (Miksis-Olds and Miller, 2006). Sites were selected based on manatee presence identified from aerial survey data from 2000 to 2003. Aerial survey data were obtained from Mote Marine Laboratory, and details of the surveys, habitat definitions, site selection criteria, and site details are provided in Miksis-Olds and Miller (2006) and Gannon et al. (2007). Percent usage for each site is defined as the percentage of surveys that manatees were observed in a specific site over the 3-year survey period. In summary, percent usage in the selected grassbeds ranged from 5% to 79%. Dredged habitat usage ranged from 5% to 40% over the same period. The pattern of site usage was not observed to change considerably from year to year over the period of 2000–2003, so it was assumed that the same pattern continued in 2004.

The quality of each grassbed site in terms of manatee foraging was characterized as part of a previous study during the same time period (Miksis-Olds and Miller, 2006). Total seagrass coverage and individual seagrass species [turtle grass (*Thalassia testudinum*), manatee grass (*Syringodium filiforme*), and shoal grass (*Halodule wrightii*)] coverage varied widely among the 13 seagrass habitats sampled. Miksis-Olds (2006) found no significant correlation between manatee grassbed usage and either total grass coverage, individual species coverage, or aerial pattern. For the parameters measured, usage was not shown to vary as a function of seagrass habitat quality in relation to density. This does not necessarily reflect patterns of usage in relation to other unmeasured parameters of seagrass quality such as percentage of cropped shoots, shoot age, plant weight, etc.

A. 2003 acoustic recordings

During the 2003 season, 5 to 10 min recordings were made at each site with a HTI-99-HF hydrophone with built-in preamplifier. This hydrophone had an operational frequency range of 2–125 000 Hz and a sensitivity of
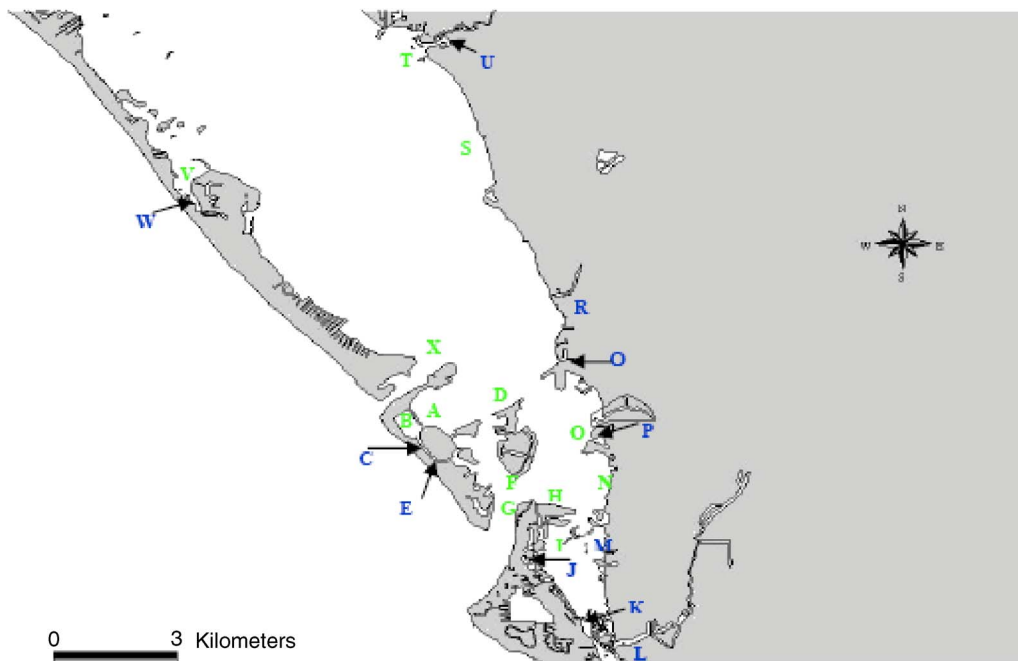
FIG. 1. (Color online) Site locations in Sarasota Bay.

−178 dB *re* 1 V/μPa. The analog signal was transferred to a Dell Inspiron 8110 via a National Instruments PCMCIA DAQ Card-6062E and digitized using the Chickadee Multi-channel Recorder Version 1.9b software program. Recording system A refers to the entire recording chain composed of the components described above (hydrophone, DAQ card, and computer). All recordings were made at a sampling rate of 200 kHz.

Acoustic site sampling was conducted from April through September in 2003. Acoustic recordings of the 24 selected manatee habitats were made systematically one to two times per week, including weekends and holidays. All identified sites were sampled in succession over the period of a few hours on the same day. The time of day for sampling was rotated among three time periods: (1) morning (07:00–10:30), (2) noon (10:30–14:00), and (3) afternoon (14:00–17:30). The order of site recordings was based on geographical location. Each site was assigned a letter in a counter-clockwise direction around the study area. The site at which the sampling regime began each week was selected randomly. Efforts were made to sample each site during each of the three time periods at least once per month (Fig 2). This

was achieved during the morning and noon time periods, but fewer recordings were made during the afternoon time period due to late afternoon thunderstorms throughout the season. The above average sampling during the noon hours in sites A, B, and C is a by-product of where animals were most often encountered. Sites B and C were part of a manatee sanctuary, and site A was adjacent to the sanctuary. Noise recordings were taken during all animal sightings and focal animal follows performed as part of a collaborative field effort. These recordings were made in addition to days devoted strictly to acoustic site sampling. In total, 395 recordings were made across the 24 sites in 2003.

B. 2004 acoustic recordings

In 2004, a different method for recording ambient noise was implemented. The goal of this sampling regime was to record noise levels in a single site over a longer consecutive period of time compared to 2003 recordings. This sampling was designed to provide better sampling of diurnal noise patterns. A passive acoustic listening (PAL) buoy was deployed for 3–4 days at each site. Each deployment included a minimum of two weekdays and one weekend day. This system consisted of a bottom mounted hydrophone system with a low-noise broadband hydrophone (100 Hz to 50 kHz), electronic filter and two-stage amplifier, TT8 computer processor with 100 kHz A/D sampler, 50 Mbyte memory card, and 51 amp-hour battery package (Nystuen, 2004). This system is referred to as recording system B.

The PAL buoy was programmed to power on and sample at 10-min intervals. The 10-min sampling interval was selected in order to preserve the battery pack over the field season and to conserve data storage space. Within each sampling interval, four 10-ms time series were sampled at 5-s intervals over a 20-s time period. This sampling protocol was
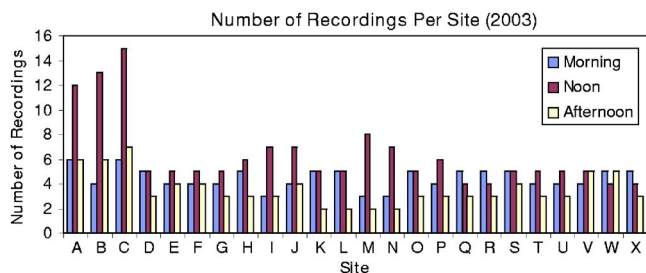
FIG. 2. (Color online) Number of recordings made at each site during each time period of the day. Grassbed sites are A, B, D, F, G, H, I, N, O, S, T, V, and X. Dredged habitat sites are C, E, J, K, L, M, P, Q, R, U, and W.

TABLE I. Average one-third octave noise levels in grassbeds (GB) ($n=13$) and dredged (DB) habitats ($n=11$) as a function of time of day in 2003, morning (7:00–10:30), noon hours (10:30–14:00), afternoon (14:00–17:30). The range of average site levels is presented in parentheses.

Time of day	Habitat	Frequency								
		250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz	32 kHz	64 kHz
Morning	GB	68 (52–80)	68 (49–80)	68 (54–80)	67 (47–79)	66 (47–79)	67 (47–79)	68 (47–79)	68 (47–79)	67 (47–79)
	DB	49 (37–67)	50 (38–67)	50 (37–67)	51 (37–69)	58 (37–83)	64 (37–91)	65 (37–91)	63 (38–86)	59 (39–82)
Noon	GB	74 (61–96)	74 (61–95)	74 (60–95)	73 (56–94)	73 (56–94)	73 (57–94)	74 (57–95)	73 (56–95)	73 (56–95)
	DB	55 (47–74)	55 (46–74)	55 (44–73)	55 (43–73)	60 (43–78)	65 (43–87)	66 (44–85)	64 (44–80)	62 (44–76)
Afternoon	GB	79 (61–92)	80 (64–91)	79 (66–91)	78 (61–91)	77 (58–91)	77 (57–91)	77 (58–91)	77 (58–91)	77 (58–91)
	DB	57 (46–79)	57 (47–79)	58 (48–79)	57 (46–78)	58 (45–79)	63 (45–80)	63 (46–81)	61 (46–80)	60 (46–79)

internally hardwired. A 1024-point fast Fourier transform (FFT) (0–50 kHz) was then performed on the time series to generate power spectra. The power spectral density (psd) curves for each 10-ms sample were then processed to identify sound sources present by comparison to stored psd curves of known sources. Data were stored as 200-Hz band averages in 0.1-dB resolution from 0 to 3 kHz, and 1-kHz band averages thereafter to conserve file space. If any of the four individual psd curves did not trigger a modified sampling regime, the spectra were averaged, stored as a single spectrum for the sampling period, and the instrument returned to "sleep" mode. However, if one of two criteria were met during the analysis of the four psd curves, the sampling regime would be modified to a 2-min sampling interval instead of 10. The first criterion was designed to reject transients. If one of the four spectra differed by more than 20 dB, indicating a transient source, the instrument was programmed to discard the first sampling set and modify its sampling regime. If transients were detected in the three consecutive sampling sets, the four psd curves of the last set were stored and the instrument returned to "sleep" mode. This criterion selected against the inclusion of transient signals, such as clicks from snapping shrimp, in the acoustic record. The second criterion was the detection of sound sources of interest, which for this study included boat traffic and rain. A modified sampling regime was triggered if the spectra resembled previously stored spectra indicating specific sound sources, and all four spectra were stored for each sampling set. Sampling continued at the 2-min interval until the source was no longer present.

C. Noise level analysis

In 2003, noise levels for each site recording were determined for one-third octave bands at nine center frequencies: 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz, 16 kHz, 32 kHz, and 64 kHz. A one-third octave bandpass filter with 30-dB side lobes was created for each of the specified frequencies. One-third octave noise levels were calculated at 4-s intervals and then averaged over the duration of the 5–10-min site recording to produce an average one-third octave noise level at a specified center frequency at a particular time and day. System noise for recording system A was obtained by suspending the hydrophone in-air and recording in a quiet room. All gain settings were identical to those used in

the field. The electronic noise floor of recording system A was at least 11 dB below the minimum recorded ambient noise level for all one-third octave bands.

Noise levels from each 2004 deployment were calculated for one-third octave bands at eight center frequencies. The stored psd center frequencies sampled with the PAL buoy were not at precise octave intervals, so the center frequencies closest to those analyzed in 2003 were selected: 300 Hz, 500 Hz, 1.1 kHz, 2.1 kHz, 4.6 kHz, 8.5 kHz, 16.3 kHz, and 31.9 kHz. For all frequencies, the analyzed bandwidth included the center frequencies measured in 2003. An average noise level within each of four time periods was calculated: (1) morning (07:00–10:30), (2) noon (10:30–14:00), (3) afternoon (14:00–17:30), and (4) night (17:30–24:00). The time periods were consistent with those used in 2003, with the addition of a fourth nighttime period. The electronic noise floor of recording system B was at least 8 dB below the minimum recorded ambient noise level for all one-third octave bands.

III. RESULTS

A. 2003 ambient noise levels

Ambient noise was analyzed in nine one-third octave bands spanning 250 Hz to 64 kHz during three time periods of the day. Analyses were performed to examine patterns both within and across the two habitat types. Overall noise patterns in the Sarasota Bay area differed between grassbed and dredged habitats (Table I). A repeated measure two-factor (habitat type × time of day) multivariate ANOVA showed that there was a significant habitat interaction for all frequencies except 16 kHz at the 5% significance level (16 kHz: $F=4.13$, $p=0.054$). Grassbeds were significantly louder than dredged habitats (average 14 dB louder across observed one-third octave bands). There was also a significant time of day interaction within habitat type, but not across habitat types. Unplanned, or *posthoc*, multiple comparisons using the least significant difference showed that for all frequencies, noise in the grassbeds was significantly louder in the noon and afternoon compared to morning time periods. In the dredged habitats, noise in the morning was significantly lower than noise in the afternoon for frequencies below 2 kHz. Above 2 kHz, noise in the morning tended to be the loudest.

Noise levels also differed significantly across the two habitat types, but only at particular times of day (Fig. 3 and Table I). *Posthoc* multiple comparisons of the repeated measure two-factor, multivariate ANOVA showed a significant habitat×time of day interaction for frequencies 4 kHz and higher $(3.93 < F < 5.49, \ 0.03 < p < 0.007)$. The emerging pattern was that for frequencies 4 kHz and above, a significant difference existed between the average noise level in the grassbeds and dredged habitats in the afternoon, with the grassbeds having a greater average noise level in the late afternoon. For frequencies below 4 kHz, there was no significant habitat×time of day interaction.

Regression analyses performed within each habitat type and each frequency band showed a significant correlation between usage by manatees and one-third octave band noise in grassbeds at all frequencies during the morning and afternoon time periods (Fig. 4 and Table II). Sites that were used more heavily by manatees tended to have lower levels of noise. For all significant regressions, *R*-squared values ranged from 0.41 to 0.49 in the morning and 0.56 to 0.71 in the afternoon. No significant relationship was observed for grassbeds during the noon hours or for the dredged habitats during any time period.

B. 2004 ambient noise levels

Ambient noise was analyzed at eight frequencies spanning 300 Hz to 32 kHz during four time periods of the day to test for differences within and across grassbed and dredged habitats. Results from this sampling regime showed overall noise patterns in the Sarasota Bay area differed slightly between grassbed habitats and dredged habitats (Table III). A repeated measure two-factor (habitat type ×time of day) multivariate ANOVA showed no habitat interaction across time periods of the day, so on average, grassbeds and dredged habitats had similar noise levels. However, there was a significant habitat type×time of day interaction for the lowest two frequencies measured: 250 and 500 Hz $(3.33 < F < 3.99, \ 0.01 < 0.02)$. For these two cases, noise in the grassbeds was louder than in the dredged habitats during the morning.

Regression analyses were performed within each habitat type and at each frequency in order to determine if ambient noise level was correlated with manatee usage. Results showed a correlation between usage and noise in grassbeds at frequencies below 1 kHz during the morning and afternoon time periods (Table IV). Correlations were also present in the grassbeds at noon for the lowest frequency and in the dredged basins at frequencies of 500 Hz and below in the morning and afternoon time periods. Sites that were used more heavily by manatees tended to have lower levels of noise. For all significant regressions at the 95% significance level, *R*-squared values ranged from 0.60 to 0.70 in the morning and 0.60 to 0.63 in the afternoon. No significant relationship at the 95% significance level was observed in either habitat at night or in the dredged habitats during the noon hours.

Results from 2003 and 2004 indicate that manatee usage is correlated with noise level in the grassbeds for the morn-
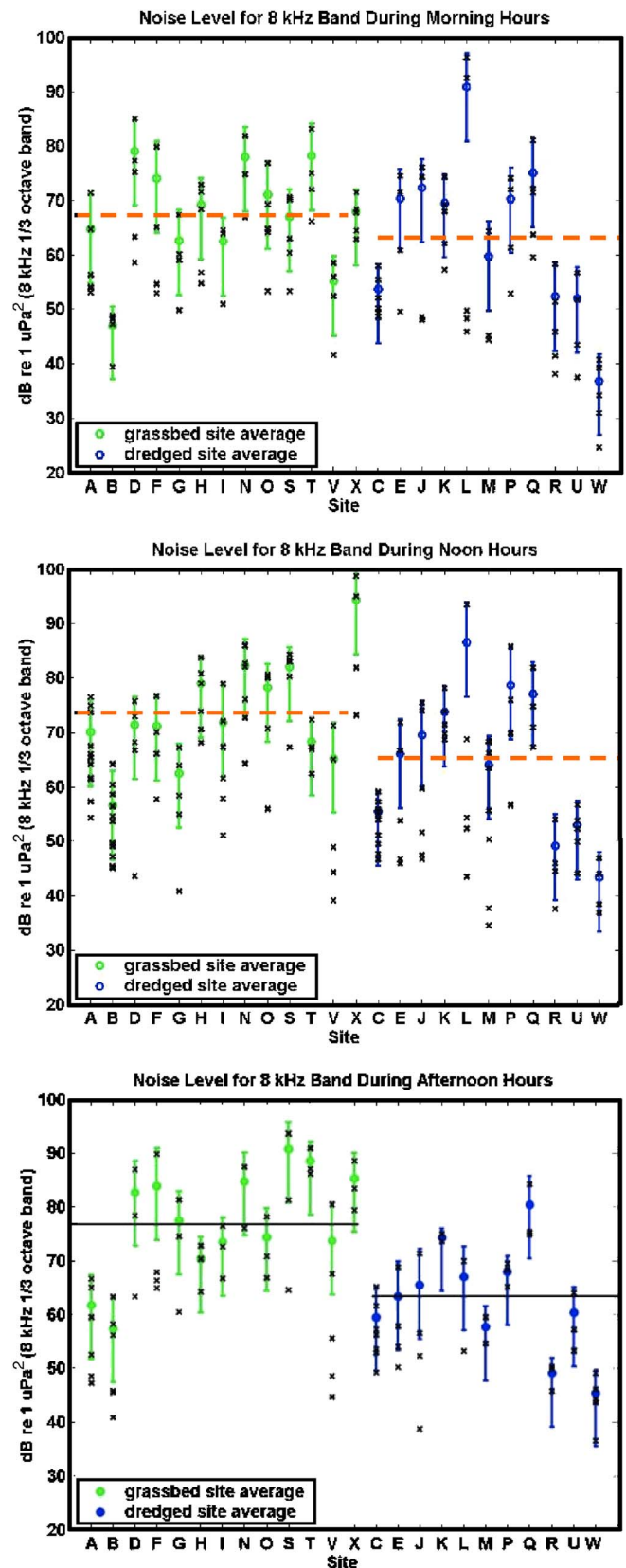
FIG. 3. (Color online) Noise levels in grassbeds and dredged habitats reported in one-third octave bands for 8 kHz center frequency. Top panel shows noise during the morning (7:00–10:30), middle panel shows noise during the noon hours (10:30–14:00), and bottom panel shows noise during the afternoon hours (14:00–17:30). Each circle represents the average level in each site with corresponding error bars. Solid black lines indicate a significant difference in means between grassbed sites and dredged habitats. Dashed red lines represent means that did not differ significantly.
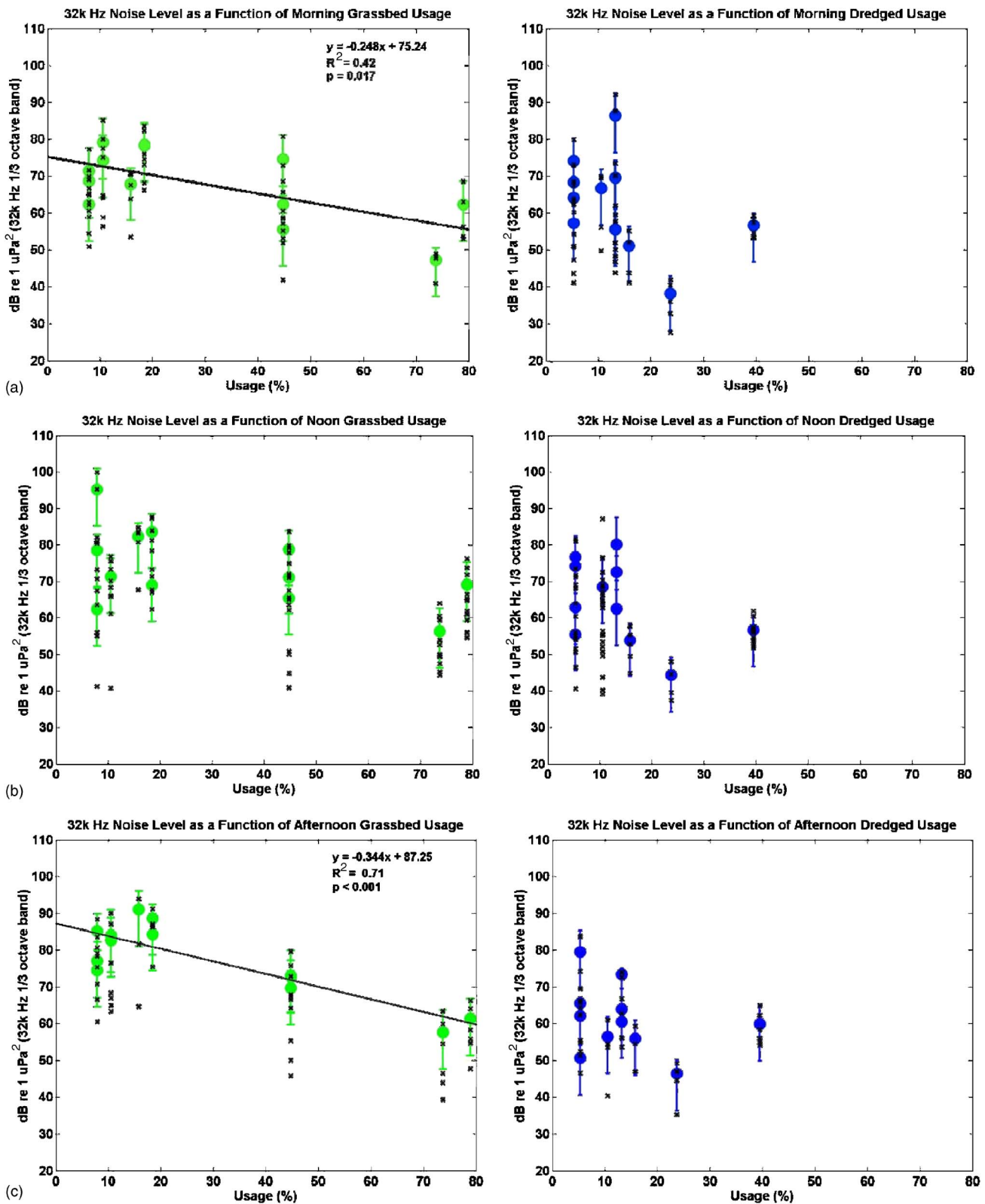
FIG. 4. (Color online) The 32 kHz noise level during morning (a), noon (b), and afternoon (c) as a function of manatee site usage in grassbeds and dredged habitats. Each circle represents the average level in each site with corresponding error bars. The *x* symbols represent individual measurements. The solid black line indicates a significant relationship at the 95% significance level.

ing and afternoon time periods. The results of both sampling methods overlap for the lowest two frequency categories (Tables II and IV). Areas that manatees used most tended to be quieter.

C. 2003 versus 2004

An argument can be made that the observed differences in noise levels and corresponding usage correlations between

TABLE II. Regression analysis *p* values for ambient noise levels and usage comparisons at specified frequencies during three times of day in 2003. Italicized values show significant relationships.

Frequency	Morning	Noon	Afternoon
	Grassbeds		
250 Hz	*0.008*	0.1	*<0.001*
500 Hz	*0.019*	0.14	*0.001*
1 kHz	*0.007*	0.12	*0.003*
2 kHz	*0.006*	0.1	*<0.001*
4 kHz	*0.008*	0.1	*<0.001*
8 kHz	*0.017*	0.11	*<0.001*
16 kHz	*0.017*	0.11	*<0.001*
32 kHz	*0.017*	0.09	*<0.001*
64 kHz	*0.008*	0.09	*<0.001*
	Dredged habitats		
250 Hz	0.31	0.31	0.78
500 Hz	0.41	0.41	0.26
1 kHz	0.24	0.31	0.54
2 kHz	0.23	0.23	0.65
4 kHz	0.15	0.18	0.54
8 kHz	0.15	0.18	0.3
16 kHz	0.19	0.16	0.33
32 kHz	0.19	0.14	0.37
64 kHz	0.25	0.17	0.47

TABLE IV. Regression analysis *p* values for ambient noise levels and usage comparisons at specified frequencies during four times of day in 2004. Italicized values show significant relationships.

Frequency	Morning	Noon	Afternoon	Night
	Grassbeds			
250 Hz	*<0.001*	0.08	0.03	0.27
500 Hz	0.01	0.13	0.05	0.94
1 kHz	0.05	0.12	0.19	0.88
2 kHz	0.41	0.5	0.43	0.53
4 kHz	0.51	0.39	0.44	0.47
8 kHz	0.4	0.54	0.58	0.44
16 kHz	0.23	0.51	0.46	0.3
32 kHz	0.24	0.52	0.62	0.43
	Dredged habitats			
250 Hz	*0.01*	0.06	*0.02*	0.07
500 Hz	*0.03*	0.12	*0.04*	0.14
1 kHz	0.37	0.56	0.21	0.16
2 kHz	0.64	0.25	0.16	0.2
4 kHz	0.52	0.22	0.11	0.24
8 kHz	0.59	0.14	0.16	0.27
16 kHz	0.26	0.19	0.14	0.18
32 kHz	0.46	0.28	0.3	0.3

the two years were due to changes in noise characteristics instead of the different sampling techniques. Noise recordings obtained during all animal observations in 2004 were made with the same recording system (recording system A) used in 2003. Noise recordings utilizing recording system A in 2004 were limited to sites in which animals were commonly encountered (sites A, B, and C). The ANOVA results showed that noise levels in these sites did not differ significantly between 2003 and 2004 (Table V). Based on these results, it can be assumed that the overall noise levels in the bay did not change significantly from 2003 to 2004.

D. Boat noise

The question of whether boats play a dominant role in the observed noise patterns can begin to be answered by analyzing the number of boats present in the vicinity of each site during the 2003 acoustic recordings. The number of boats passing within a 1-km radius of the research vessel was documented during each acoustic recording in 2003. A 1-km radius was chosen because manatees have been shown to respond to approaching boats up to 1 km away (Nowacek *et al.*, 2002). A set of single classification ANOVAs showed there was no significant difference in the average number of boats per site as a function of time of day in grassbeds or dredged habitat. However, there was a correlation at the 95% significance level between manatee grassbed usage and the average number of boats passing/5–min time period during the morning hours ($R^2=0.409$, $p=0.018$) (Fig. 5). A detailed analysis of the corresponding noise levels showed a positive correlation between morning noise level and the average number of boats in grassbeds (Fig. 6). Linear regression showed a significant increase in grassbed noise level with an increase in boat presence during the morning hours for all frequencies examined (Table VI). There was no correlation between number of boats and manatee usage of grassbed sites during the noon or afternoon hours or in dredged habitats at any time of day.

TABLE III. Average noise levels in grassbeds and dredged habitats as a function of time of day obtained from PAL measurements in 2004. All noise level values are presented in dB *re* 1 μPa. Ranges are presented in parentheses.

Time of day	Habitat	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz	32 kHz
Morning	GB	67 (53–79)	62 (52–72)	57 (52–66)	54 (46–71)	52 (37–72)	49 (34–69)	45 (33–64)	41 (31–58)
	DB	65 (57–76)	60 (53–71)	58 (51–69)	55 (41–73)	54 (40–71)	49 (37–64)	46 (33–58)	42 (31–52)
Noon	GB	69 (56–81)	63 (57–74)	57 (44–64)	55 (43–68)	53 (43–66)	48 (41–68)	45 (36–62)	41 (32–57)
	DB	67 (59–81)	62 (55–74)	58 (52–73)	55 (42–72)	54 (41–70)	49 (38–63)	47 (39–55)	43 (34–52)
Afternoon	GB	68 (54–82)	62 (51–76)	55 (44–63)	54 (46–66)	51 (45–66)	47 (38–64)	44 (36–53)	40 (33–56)
	DB	67 (57–81)	62 (54–73)	58 (51–68)	55 (44–68)	53 (40–65)	49 (38–65)	47 (35–66)	43 (33–65)
Night	GB	68 (56–81)	63 (54–72)	55 (46–62)	51 (42–65)	48 (37–67)	45 (32–65)	44 (32–63)	41 (31–51)
	DB	68 (58–79)	62 (55–70)	60 (54–63)	57 (45–65)	54 (40–65)	50 (37–60)	47 (35–56)	43 (33–50)

TABLE V. Average 1/3-octave band noise level comparisons for 2003 and 2004.

		250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz	32 kHz	64 kHz
Site A	2003	66	70	68	65	64	66	65	64	64
	2004	63	63	62	61	61	62	62	62	62
Site B	2003	58	58	60	55	53	54	54	54	54
	2004	54	55	53	53	52	53	53	53	53
Site C	2003	52	54	51	51	52	56	58	58	57
	2004	53	53	52	53	53	55	58	57	56

TABLE VI. Linear regression statistics indicating a significant positive correlation between boat presence and noise level.

Frequency	R^2	F	p value
250 Hz	0.41	7.80	0.017
500 Hz	0.32	5.29	0.041
1 kHz	0.36	6.22	0.030
2 kHz	0.37	6.36	0.028
4 kHz	0.36	6.09	0.031
8 kHz	0.36	6.08	0.031
16 kHz	0.37	6.54	0.027
32 kHz	0.43	8.28	0.015
64 kHz	0.35	7.42	0.020

IV. DISCUSSION

Manatees live in an acoustic environment that is highly variable in both space and time. Noise intensity level can change by orders of magnitude in a span of seconds. In general, grassbeds tended to be louder than dredged habitats due to the loud broadband noise produced by snapping shrimp (*Alpheus and Synalpheus sp*), which becomes stronger with decreasing depth (Camp *et al.*, 1998; Richardson *et al.*, 1995, Miksis-Olds, 2006). The noise level patterns, however, differed greatly between the two years due to the different sampling methods. In 2003, recordings were made with a single hydrophone and processed at a later date. Processing of the 2003 data included all sounds present in the environment without any special weighting or selection. Noise produced by snapping shrimp dominated the shallow habitat noise recordings in 2003; furthermore, the 2003 noise level analysis did not distinguish between the confounding factors of broadband shrimp noise and lower frequency anthropogenic noise. Noise recordings in 2004 were also made from a single hydrophone, but initial processing of the noise spectrum was done in real time. With this processing protocol, transients such as broadband signals produced by snapping shrimp were selected against, so the resulting noise recordings more accurately reflected patterns associated with low frequency anthropogenic noise as opposed to the loud biological signal of snapping shrimp.

This study benefited from the implementation of two different noise sampling techniques that allowed for the separation of confounding factors associated with the recordings in 2003. Filtering the transient signals out of the acoustic record in the 2004 data set essentially removed the broadband shrimp sounds, which allowed for the identification of patterns associated with lower energy, more narrow-band anthropogenic noise. Based on the analysis of habitat use and noise recordings in 2003, manatees select grassbed sites that have lower noise levels across a wide range of frequencies. The relationship between usage and noise, however, was only present during the morning and late afternoon time periods. Although snapping shrimp noise is constantly present, these time periods are associated with both sunrise and sunset, which happens to be linked with the small diurnal increase in activity that snapping shrimp exhibit at night (Albers, 1965; Au and Banks, 1998). The 2003 regression results suggest that manatees select the quietest grassbed sites during the peak times of snapping shrimp activity, which implies that manatees select grassbeds with less snapping shrimp noise. The 2004 recordings reduced the presence of the snapping shrimp noise, although they did exhibit a slight diurnal variation associated with this biologic noise source. The analysis of the 2004 data indicated that manatees selected grassbed sites that have less low frequency noise, which is most likely attributed to anthropogenic activity. The question of whether shrimp noise or anthropogenic noise is the major force behind the observed manatee distribution patterns cannot be answered conclusively with the current data, and the possibility remains that the presence or absence of snapping shrimp correlates with some aspect of habitat quality that was not measured. However, it is clear that noise is one factor that correlates with manatee habitat selection.

The current study builds upon previous transmission loss measurements conducted in the same sites of Sarasota, FL over the same time period. Miksis-Olds and Miller (2006) showed that grassbeds used most often by manatees had high levels of transmission loss above 2 kHz. However, it was noted that information pertaining to environmental noise levels in specific habitats is needed in order to more fully understand questions pertaining to manatee habitat selection and signal detection. By combining information about habitat usage in relation to transmission loss characteristics with daily noise patterns (both biologic and anthropogenic), an interesting two-part picture starts to emerge. Manatee usage of grassbed habitats was highly correlated with broadband noise, low frequency noise, and high frequency transmission loss. High-use grassbeds were areas of high transmission loss (frequencies >2 kHz) and low broadband and low frequency noise (frequencies <1 kHz). This creates a high-use environment where noise above 2 kHz from sources outside the

FIG. 5. (Color online) Manatee site usage as a function of boat presence in grassbeds.

1 kHz Noise Level as a Function of Number of Boats

$y = 0.038x - 1.7992$
$R^2 = 0.3612$
$p = 0.029$

Avg # boats/ 5 min time period (y-axis)

dB re 1 μPa^2 (8 kHz 1/3 octave band) (x-axis)
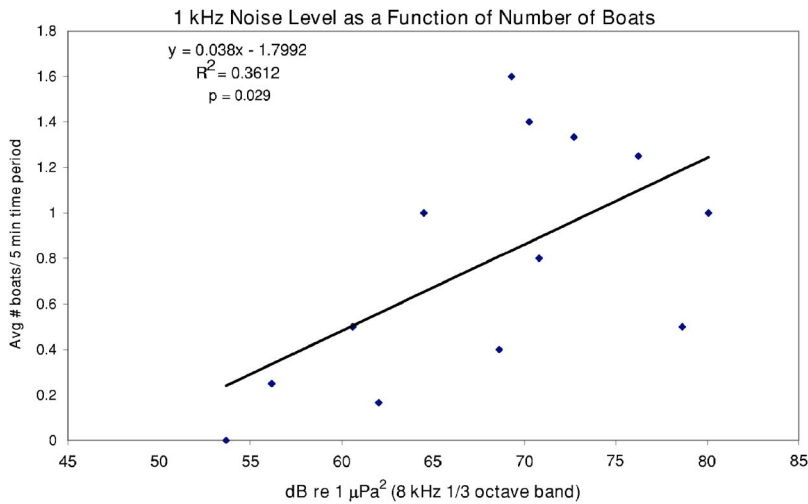
FIG. 6. (Color online) The 1 kHz noise level as a function of boat presence in grassbeds during the morning hours.

grassbed is attenuated quickly compared to low-use sites. This also happens to be the range of most efficient sound propagation inside the grassbed habitat, and the dominant frequencies of manatee vocalizations (Nowacek *et al.*, 2003). For frequencies below 2 kHz, transmission loss is not correlated with usage, but low frequency noise is, especially during the morning hours. Ultimately, the grassbeds that manatees selected were those that were most quiet due to lower noise below 1 kHz and higher transmission loss above 2 kHz. The selection of dredged habitats was not correlated with either noise level or transmission loss, but the dredged habitats used most often by manatees were in close proximity to high-use grassbeds.

The next logical question is what feature is more influential in driving the manatee grassbed usage, noise or habitat quality? Analysis of the seagrass coverage and species composition indicated no correlation between quality and grassbed usage (Miksis-Olds and Miller, 2006). This suggests that noise and associated transmission loss characteristics may play a more dominant role in habitat selection, compared to the limited grassbed quality parameters investigated. This finding contributes to the interpretation of manatee habitat selection and presents a need for clear-cut hypotheses to be experimentally tested in the field. For example, will increasing noise in a high-use grassbed alter use over the course of a week, month, or year? Will decreasing noise in a low-use area increase grassbed use? These questions need to be answered for effective conservation management, for they will be instrumental in directing future management plans.

Time of day also has an effect on how manatees are using their habitat. High-use areas have less low frequency noise at a time when overall noise levels are significantly increasing throughout the bay during the daily night-to-morning transition. It is during this transition period that boat use also increases. There is a concentrated increase of boat use at daybreak, possibly associated with fisherman going out to fish. No concentrated return was observed in the afternoon, as return times appeared widespread possibly due to weather, fishing success, etc. Grassbed usage was negatively correlated with the concentrated boat presence in the morning hours. This suggests that morning boat presence and its associated noise may play a dominant role in grassbed usage

on a daily time scale. Other factors to consider are tidal cycle, current, and the change of noise level per unit time.

As Sarasota Bay manatees predominantly use grassbeds to feed (Koelsch, 1997), it can be extrapolated that the presence of boats in the morning and their associated noise may affect manatee foraging behavior. This result is not unique to manatees, as this pattern has also been observed in another endangered species, the wintering bald eagle (*Haliaeetus leucocephalus*). Eagle numbers on the Skagit River Bald Eagle Natural Area (SRBENA) in northwestern Washington were negatively correlated with daily boat traffic, and feeding declined exponentially with increased boating activity (Stalmaster and Kaider, 1998). However, early morning boat traffic was most disruptive to eagle feeding behavior. Eagles took longer to return to foraging sites during morning disturbances than later in the day.

Masking of important biological signals is yet another impact of rising ambient noise levels and transient sources. On average, grassbeds were 14 dB louder than dredged habitats. The average level of ambient noise in high and low use grassbeds, as a function of frequency, differed as much as 30 dB. With ambient noise levels reaching over 90 dB *re* 1 μPa within multiple one-third octave bands in some grassbeds, it is not unreasonable to estimate that the time at which a manatee first detects a transient noise source, such as an approaching watercraft, could be vastly decreased in high noise conditions. For example, a mere 4-s delay in detection results in the potential loss of approximately 25 m when swimming at top speed [estimated using a top swim speed of 6.25 m/s (Hartman, 1979)].

Masking also reduces the range of effective vocal communication between manatees. Estimates of manatee vocalization source levels range from 106 to 115 dB root mean square (rms) *re* 1 μPa at 1 m (Nowacek *et al.*, 2003; Phillips *et al.*, 2004). In addition, Schevill and Watkins (1965) reported that the vocalizations were not particularly loud, registering only 10–12 dB above the background noise at 3–4 m in a vegetation-choked canal. Increases in ambient noise levels on the order of 10–12 dB are not uncommon, but could have drastic repercussions for the manatee in terms of effective range of communication and the energy required

to maintain effective communication if they respond to noise by increasing their source level.

ACKNOWLEDGMENTS

Albers, V. M. (**1965**). *Underwater Acoustics Handbook—II* (The Pennsylvania State U.P., University Park, PA).

Au, W. W. L., and Banks, K. (**1998**). "The acoustics of the snapping shrimp *Synalpheus parneomeris* in Kaneohe Bay," J. Acoust. Soc. Am. **103**, 41–47.

Camp, D. K., Lyons, W. G., and Perkins, T. H. (**1998**). "Checklists of Selected Marine Invertebrates of Florida," Report FMRI Technical Report TR-3, Florida Department of Environmental Protections.

Etter, P. C. (**1996**). *Underwater Acoustic Modeling* (E & FN SPON, London).

Gannon, J. G., Scolardi, K. M., Reynolds, J. E., III, Koelsch, J. K., and Kessenich, T. J. (**2007**). "Habitat selection by manatees in Sarasota Bay, Florida," Marine Mammal Sci., **23**, 133–143.

Koelsch, J. (**1997**). "The seasonal occurrence and ecology of Florida manatees (*Trichechus manatus latirostris*) in coastal waters near Sarasota, FL," M.S. thesis, University of Southern Florida, Tampa.

Kryter, K. D. (**1985**). *The Effects of Noise on Man*, 2nd ed. (Academic, Orlando, FL).

Marine Mammal Commission. (**1986**). "Habitat protection needs for the subpopulation of West Indian manatees in the Crystal River area of northwest Florida," Document No. PB86-200250 (National Technical Information Service, Silver Springs, MD).

Miksis-Olds, J. L. (**2006**). "Manatee response to environmental noise," Ph.D. dissertation, University of Rhode Island, Narragansett.

Miksis-Olds, J. L., and Miller, J. H. (**2006**). "Transmission loss in manatee habitats," J. Acoust. Soc. Am. **120**, 2320–2327.

Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations (L)," J. Acoust. Soc. Am. **114**(1), 66–69.

Nowacek, S. M., Nowacek, D. P., Johnson, M. P., Shorter, K. A., Powell, J. A., and Wells, R. S. (**2002**). "Manatee behavioral responses to vessel approaches: Results of digital acoustic logger tagging of manatees in Belize," Mote Marine Laboratory Technical Report No. 847, Mote Marine Laboratory, Sarasota, FL.

Nystuen, J. (**2004**). "Using acoustic rain gauges to monitor underwater sound," Workshop on Ocean Ambient Noise Budgets and Long Term Monitoring: Implications for Marine Mammals, Warwick, RI.

Phillips, R., Niezrecki, C., and Beusse, D. O. (**2004**). "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**, 422–428.

Richardson, W., Greene, C., Malme, C., and Thomson, D. (**1995**). *Marine Mammals and Noise* (Academic, San Diego, CA).

Schevill, W. E., and Watkins, W. A. (**1965**). "Underwater calls of *Trichechus* (Manatee)," Nature (London) **205**, 373–374.

Sousa-Lima, R. S., Paglia, A. P., and Da Fonseca, G. A. B. (**2002**). "Signature information and individual recognition in the isolation calls of the Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia)," Anim. Behav. **63**, 301–310.

Stalmaster, M., and Kaider, J. (**1998**). "Effects of recreational activity on wintering bald eagles," Wildlife Monogr **137**, 1–46.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula, Los Altos, CA.).

Hartman, D. S. (**1979**). "Ecology and behavior of the manatee (Trichechus manatus) in Florida." American Society of Mammalogists, Special Publication 5.

Session 1aAAa

Architectural Acoustics, Noise, Engineering Acoustics, and ASA Committee on Standards: In Situ Measurements of Absorption Coefficients

Peter D'Antonio, Cochair
RPG Diffusor Systems Inc., 651C Commerce Dr., Upper Marlboro, MD 20774

Christian Nocke, Cochair
Akustikbuero Oldenburg, Katharinenstr. 10, D-26121 Oldenburg, Germany

Chair's Introduction—8:00

Invited Papers

8:05

1aAAa1. Short review on *in situ* measurement techniques of impedance or absorption. Christian Nocke (Akustikbuero Oldenburg, Katharinenstr. 10, D-26121 Oldenburg, Germany)

The paper will present a short historical review on the development of *in situ* measurement techniques of sound absorption or the acoustic surface impedance. One of the earliest setups to measure the absorption of a material *in situ* has been proposed in 1933. In 1934 a method applying short tones to separate the reflected signal from the incident signal in front of a reflecting surface has been proposed. Many more methods have been described over the years. Applications of modern MLS-based measurement equipment to deduce the absorption coefficient *in situ* were brought up in the early 1990s. An MLS-based procedure similar to an early method has been introduced as subtraction technique and is the basis of European standard ENV 1793, part 5. Most of these methods are based on the assumption of plane wave propagation. Other methods relying on spherical wave propagation approach are being reviewed. Measurement examples and applications in room acoustics will be shown for some of the methods presented and compared to each other.

8:25

1aAAa2. *In situ* measurements of the absorption coefficient in porous materials performed with parametric arrays and with standard audio instrumentation. Bernard Castagnede and Alexei Moussatov (Lab. d'Acoustique de l'Université du Maine, Ave. Olivier Messiaen, 72017 Le Mans Cedex 9, France, Bernard.Castagnede@univ-lemans.fr)

Metrology applications in the field of the characterization of sound absorbing porous materials were considered quite early by using ''parametric arrays'' of nonlinear acoustics. Some experimental work was performed in our lab with various devices having a central frequency varying from 40 kHz to 200 kHz [Castagnede *et al.*, Ultrasonics, **44** (2006)]. Reflection and transmission coefficients are available versus frequency, as well as dispersion curves. Experimental data are compared to numerical predictions in the frame of the standard poroelastic models for different porous materials. Some dedicated instrumentation has also been designed for industrial applications, e.g. for ''on-line'' and ''in-situ'' measurements of the LF coefficient of absorption over the 100 Hz–6 kHz bandwidth [Castagnede *et al.*, Applied Acoustics, **68** (2007)]. More recently, absorption coefficient measurements have been performed with the same type of configurations with standard audio instrumentation, including a simple loudspeaker and a single microphone. These data are compared here, for numerous fibrous materials, with those obtained with parametric arrays, resonance Kundt tubes, as well as with numerical predictions. The agreement between the various measurements is generally good, even better at low frequencies with the audio system, because the LF sensitivity (below 200 Hz) of the parametric devices is generally quite poor.

8:45

1aAAa3. Determination of absorption coefficients for audience seating from *in situ* measurements. Timothy E. Gulsrud and C. Walter Beamer IV (Kirkegaard Assoc., 954 Pearl St., Boulder, CO 80302)

Audience seating constitutes the majority of fixed sound absorption in most music performance venues. An understanding of the sound absorption characteristics of audience seating is therefore essential for accurately predicting reverberation times and for evaluating different chair constructions during the design process. Room acoustics measurements were recently conducted in a small concert hall immediately before and after installation of audience seating. These measurements allow the determination of absorption coefficients for audience seating for use in the Sabine equation [L.L. Beranek and T. Hidaka, J. Acoust. Soc. of Am. **104**, 3169–3177 (1998)] and in room acoustics computer modeling software. A comparison of these *in situ* absorption coefficients along with those determined from reverberation room measurements will be discussed.

3029   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3029**

1aAAa4. Experience with *in situ* measurements using Electronic and Acoustic System Evaluation and Response Analysis (EASERA). Wolfgang Ahnert, Stefan Feistel (SDA Software Design Ahnert GmbH, Arkonastrae 45-49, 13189 Berlin, Germany, info@sda.de), and Bruce Olson (SDA Software Design Ahnert GmbH, Brooklyn Park, MN)

To measure the absorption coefficients of built-in materials is quite often necessary, however, in reality the results deviate from those values measured in the reverberation chamber. Using the measurement platform EASERA our experiences are reported which we gathered during measurements in a big assembly hall. The measurement procedure will be explained and the needed software and hardware tools are described. It turned out that the quality of the reference signal and the reflected one is crucial for the quality of the obtained results. Hints for the right windowing are expounded to get usable impulse responses. Also, it is explained how the selected window limits influence the stability of the results.

1aAAa5. Aspects of direct deduction of ground impedance. Keith Attenborough (Dept. of Eng., The Univ. of Hull, Cottingham Rd., Hull HU6 7RX, UK), Shahram Taherzadeh (The Open Univ., Milton Keynes MK7 6AA, UK), Gilles Daigle (NRCC, Ottawa, Canada), and Roland Kruze (Univ. of Oldenburg, Oldenburg, Germany)

ANSI S1.18 1998 offers a method for determining ground impedance by fitting data to templates of the magnitude of level difference obtained using a point source and two vertically separated microphones with specified geometries according the ''acoustical-softness'' of the ground. The ANSI working group on ground impedance is seeking to revise the current template method for deducing ground impedance by developing a practical method, based in fitting complex level difference spectra and without assuming any particular impedance model. Both minimization and root finding techniques have been used in published laboratory experiments. The results of trial outdoor measurements in the UK, USA, Germany, and Canada are presented. Issues have arisen concerning phase matching of microphones, the signal processing instrumentation, the site and meteorological constraints, and the smoothness of the resulting data.

1aAAa6. Optimized geometry for ground impedance measurement. Roland Kruse, Volker Mellert, and Ping Rong (Oldenburg Univ., Inst. of Phys., 26111 Oldenburg, Germany, roland@aku.physik.uni-oldenburg.de)

The two-microphone method is a convenient and well-known procedure for the *in situ* determination of the surface impedance both in room acoustics and for outdoor use. In light of its possible use in a future revision of ANSI S1.18-1999, its sensitivity towards measurement errors has been investigated. Unfortunately, both measurements and simulations show that the method is very sensitive to errors, especially in the transfer function, at low frequencies, and for high surface impedances when using one of the three suggested geometries from the standard. Therefore, an improved geometry is proposed to reduce the effect of errors. However, even with this optimization, the use of the two-microphone method cannot be recommended for surfaces with a high impedance at low frequencies as the necessary precision is unacceptably high.

Contributed Paper

1aAAa7. The "effective" sound absorption of unique variable acoustic devices in Jensen Concert Hall, Pocatello, Idaho. William Chu and David Conant (McKay Conant Brook 5655 Lindero Canyon Rd., #325, Westlake Village, CA 91362)

The results of acoustical measurements taken at a new 1200 seat concert hall with variable acoustics demonstrate that significant reverberation control can be obtained with strategically located, but rather unusual, absorptive devices. These unique elements yield a change in RT60 of 1 second, with insignificant reduction in room volume and minimally exposed absorptive area. Various measures for estimating RT60 are compared, along with CATT-Acoustic modeling, to assess the best characterization of the influence of these devices, including an "effective" sound absorption coefficient.

Session 1aAAb

Architectural Acoustics: Measurements and Simulation

Bruce C. Olson, Chair

Olson Sound Design, 8717 Humboldt Ave., North, Brooklyn Park, MN 55444-1320

Chair's Introduction—10:25

Contributed Papers

10:30

1aAAb1. Segmented swept sine technique for room impulse response estimation. Joel P. Paulo, Jose L. Bento Coelho, and Carlos M. Martins (ISEL, R. Cons. Emidio Navarro, 1, 1950 Lisbon, Portugal)

The swept sine technique is very popular for the measurement of room impulse response due to its large signal-to-noise ratio (SNR) and immunity against subtle time-variance and nonlinearity of the room under test. However, the acoustical measurements in the presence of high background noise levels leads usually to unsatisfactory results. For situations of high level nonstationary noise, the mean square of the overall sequence must be minimized in order to increase the SNR. The new technique named segmented swept sine consists in exciting the room with a set of M swept sine signals with application of a dedicated weighting average method and a noise pattern recognition procedure. Each swept sine signal is segmented followed by the estimation of the mean square (MS) value of the respective segment. A weighting procedure is applied to each segment according to the MS value and the likelihood between them. The average technique is then applied. This procedure ensures that the resulting swept sine signal has the highest SNR value. Several examples are presented to compare the standard swept sine and the segmented swept sine techniques, giving advantages and disadvantages of each technique.

10:45

1aAAb2. Swept sine-based technique for minimization of annoyance using a perceptual model. Joel P. Paulo, Jose L. Bento Coelho, and Carlos R. Martins (ISEL, R. Cons. Emidio Navarro, 1 1950 Lisbon, Portugal)

The accurate estimation of the acoustical parameters of a space to be used by people, such as theaters, concert halls, conference rooms, sport stadiums, and other public areas, implies that the measurements should be evaluated in the presence of an audience. However, for reasons of annoyance, people are usually avoided and a correction factor related to the effective absorption of the audience area is applied to the results only. This procedure does not take into account the variability of some parameters such as the relative humidity during the event, which influences considerably the magnitude frequency response. The swept sine technique is very attractive in acoustical measurements due to the high signal-to-noise ratios (SNR) and robustness against nonlinearity of the measurement chain and time variance. The use of some convenient music tracks in agreement with the perceptual masking effect minimizes the annoyance and increases the SNR simultaneously. The sinusoidal synthesis algorithm is applied to some parts of the music. After questioning a set of persons, some results based on an annoyance indicator are presented giving advantages and disadvantages of this method.

11:00

1aAAb3. Evaluation of speech intelligibility from binaural room-acoustics measurements. Dorea Ruggles, Ning Xiang (School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180, ruggld@rpi.edu), and Wolfgang Ahnert (Ahnert Feistel Media Group, Berlin, Germany)

Efficient methods of quantifying speech intelligibility are needed for designing and understanding functional architectural spaces. All current measures of speech intelligibility are monophonic, which excludes consideration of important binaural aspects of human hearing, including dereverberation and decoloration. Acquisition of binaural data has increasingly become common practice, using dummy heads or in-ear microphones, making the development of binaural intelligibility measures especially timely. The need for such a measure is illustrated by calculations of speech transmission index (STI) with a single energy-based impulse obtained through the combination of binaural data channels. Each method of combination produces significantly different STI values that illustrate the impact of orientation and location on intelligibility calculations. The relationship between such estimation variation and subjective experience must be studied to determine the research direction for a much needed, uniquely binaural speech intelligibility measure.

11:15

1aAAb4. Assessing the speech intelligibility index in Johns Hopkins Hospital. Philip Kwon, Ilene Busch-Vishniac, and James West (Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

Over the last few years we have gathered considerable amounts of data on the sound pressure levels present in Johns Hopkins Hospital in Baltimore, MD. These data provide information on levels versus location, frequency, and time of day. Using these data we have determined the SII in various locations of the hospital. Speech intelligibility is a key issue in hospitals because the vast majority of communication takes place orally. Further, the vocabulary used in hospitals includes a significant set of unusual words such as pharmaceutical names that are easily confused. Thus, an SII higher than required in an office environment is needed for clear speech communication in a hospital. Our results suggest that speech intelligibility is compromised by the noise present in the hospital.

11:30

1aAAb5. Spatial variation of acoustic energy density in lightly damped rectangular enclosures. Buye Xu and Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

Since the 1990's, acoustic energy density (ED) has been investigated for applications such as active noise control and sound equalization problems. Thus far, the developments have generally relied on the reduced spatial variation of ED. Previous studies on the spatial variation of ED in lightly damped rectangular enclosures were limited to the low frequency range where only a small number of normal modes are excited, and the relatively high frequency range where statistical methods can be applied. In this work, the properties of ED in mid-frequency range have been

3031   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3031**

studied, based on a computer simulation model. Different conditions for the enclosure and the sound source have been studied. The results show good consistency with previous studies in the overlap ranges of frequency. [Work supported by NSF Grant CMS-0409319.]

11:45

1aAAb6. Acoustic absorbers made of micro-perforated stretched ceilings and other materials. Christian Nocke, Catja Hilge (Akustikbüro Oldenburg, Katharinenstr. 10, D-26121 Oldenburg, Germany), and Jean-Marc Scherrer (BARRISOL Normalu S.A.S, F-68680 Kembs, France)

The manufacturer of stretched ceiling systems celebrates its 40th birthday in 2007. In 2001 a nearly invisible micro-perforation was introduced to the stretched material, making it highly sound absorptive. Over the last years different setups made of micro-perforated layers, porous materials, as well as plate resonators have been investigated. As a latest addition combinations of nonwoven materials with micro-perforated layers have been experimentally investigated. In this contribution measured sound absorption coefficients of various setups with micro-perforated stretched foils and different other acoustic materials will be presented. For these assemblies no closed calculation model exists so far.

MONDAY MORNING, 4 JUNE 2007   CANYON ROOM B/C, 8:00 TO 10:30 A.M.

Session 1aAOa

Acoustical Oceanography and Physical Acoustics: Acoustics of Bubble Clouds

Grant B. Deane, Chair

Univ. of California, San Diego, Marine Physical Lab., La Jolla, CA 92093-0238

Invited Papers

8:00

1aAOa1. Cloud cavitation. Christopher Brennen (California Inst. of Technol., Pasadena, CA 91125)

In many cavitating liquid flows, when the number and concentration of the bubbles exceeds some critical level, the flow becomes unsteady and large clouds of cavitating bubbles are periodically formed and then collapse when convected into regions of higher pressure. This phenomenon is known as cloud cavitation and when it occurs it is almost always associated with a substantial increase in the cavitation noise and the potential for material damage associated with the cavitation. These increases represent serious problems in devices as disparate as marine propellers, cavitating pumps, and artificial heart valves. This lecture will present examples of the phenomenon and review recent advances in our understanding of the dynamics and acoustics of cloud cavitation. Both analyses of these complex multiphase flows and experimental observations will be used to identify the key features of the phenomenon and the parameters that influence it. [Work supported by the U.S. Office of Naval Research.]

8:25

1aAOa2. Low-frequency acoustic resonances from the axi-symmetrical bubble plume. Michael J. Buckingham, Thomas R. Hahn, and Thomas K. Berger (MPL, SIO, UCSD, 7900 Gilman Dr., La Jolla, CA 92093-0238)

Experiments were performed on the sound from an axi-symmetric, conical bubble plume formed by a vertical freshwater jet as it plunges into a freshwater pool. The fluxes of air and water entering the pool were controlled during the experiments, and, outside the plume sound was monitored over the frequency band 0.1 to 1 kHz. Up to five non-uniformly spaced peaks were observed in the acoustic spectrum, associated with the coherent, collective oscillations of the bubbles within the plume. The biphasic bubbly medium behaves as a continuum, acting as a resonant, conical cavity beneath the jet. The eigenfrequencies scale inversely as the square-root of the jet velocity and the fourth root of the air entrainment ratio. A two-component, theoretical model for the eigenfrequencies involves a fluid-dynamics argument, which shows that the sound speed increases as the square-root of depth in the plume. This is incorporated into an acoustic analysis in which the wave equation is solved analytically, taking account of the cone-like geometry of the bubble plume cavity. The theoretical frequencies of the lowest-order longitudinal modes of the bubbly cavity are found to exhibit the same inverse-fractional power-law scalings as those observed in the experiments. [Research supported by ONR.]

8:50

1aAOa3. Sound channeling through bubbly streams. Aneta Nikolovska (Res. Ctr. of Ocean Margins, MARUM, Leobener Str., Univ. of Bremen, 28359 Bremen, Germany, aneta@uni-bremen.de)

The presence of even small amount of bubbles can have a strong effect on the transmission and reduction of the sound waves through the medium in which they occur. The study presented here examines experimentally and numerically the propagation of acoustic energy in the vicinity of streams of discrete air bubbles. The acoustic energy was created naturally during the detachment of bubbles at the bottom of the stream. Previous work [A. Nikolovska, J. Acoust. Soc. Am. **115**, 2473 (2004)] has reported that the root-mean-squared pressure distribution is highly anisotropic in the bubbles vicinity. A new experimental technique has been developed to obtain time frames of the instantaneous acoustic pressure field using two hydrophones. This technique allowed coordinated

measurement of the acoustic energy distribution in the near and far field. The results show that the phase speed in the direction of the bubble stream has values considerably lower than the speed of sound in pure water. Different bubble stream configurations were investigated and it was found that this speed of propagation reduces and the quality factor of the sound pulses increases as the air volume fraction in the bubble streams is increased. [Work supported by RCOM.]

9:15–9:30 Break

Contributed Papers

9:30

1aAOa4. Experiment on acoustic wave propagation in a bubbly water: Frequency influence. Vincent Duro, Dominique Decultot, and Gerard Maze (LAUE UMR CNRS 6068, IUT, place Robert Schuman, 76610, Le Havre, France)

Commander and Prosperetti works [J. Acoust. Soc. Am. **85**, 732–746 (1989)] show that an acoustic wave, propagating in a bubbly water, is very strongly damped and its phase speed is disturbed for bubble resonance frequency neighboring. In relation to our experimental results, numerical results obtained with these analytical works are presented. Experiments are realized in a large water-filled tank. A frequency modulated burst, generated with broadband transducers of central frequency 100, 200, and 500 kHz, propagates through the bubble cloud obtained by a high pressure water jet. Our measurements show simultaneously, after the stop of jet, a strong damping of acoustic signal and a small change of its phase speed. After some minutes, this signal gets back progressively its initial amplitude and phase speed values. One can notice that the acoustic signal amplitude and phase speed are not identical in all emission frequency domain. To explain this phenomenon, we assume the existence of frequency band gaps related to the multiple scattering from the bubble cloud. [Work supported by Bassin d'Essais des Carènes (France).]

9:45

1aAOa5. The effect of depth-dependent distribution of bubbles on internal waves. Lev A. Ostrovsky (Zel Technologies/Univ. of Colorado, 325 Broadway, Boulder, CO 80305, Lev.A.Ostrovsky@noaa.gov), Roger H. J. Grimshaw, and Karima K. Khusnutdinova (Loughborough Univ., Loughborough LE11 3TU, UK)

A new type of gravity waves associated with the subsurface layers of bubbles existing ubiquitously in the ocean is considered. Due to bubble compressibility, such a layer creates a vertical density gradient. Although the density changes due to bubbles are small, the corresponding gradient can be significant, at least comparable with or even exceeding the natural density gradient due to the temperature or salinity variations. As a result, the bubbles may affect propagation of short-period internal waves or even be exclusively responsible for the internal modes in otherwise homogeneous water. These modes can be strongly localized near the surface and thus contribute to ripple modulation, affect sound propagation, etc. The dispersion equation for the bubble-affected internal waves is derived, and the corresponding modal structures are plotted. Some estimates are made and limitations of the simple model considered under real conditions are discussed.

10:00

1aAOa6. Estimation of drop size distribution from the low frequency components in the rain generated underwater acoustic noise. Mani Thundiyil (Model Eng. College, Cochin, India) and PRS Pillai (CUSAT, Cochin, India)

Drop size distribution (DSD) is an important parameter that completely characterizes the rain and rain generated acoustic energy. This paper presents a simple and efficient method for measuring the DSD of rain by analyzing the low-frequency components in the spectrum of rain generated acoustic noise underwater. A specially designed sensor assembly captures the acoustic noise produced due to the raindrop impacts on the water surface. Every individual drop produces acoustic noise comprising the low frequency components which has distinct characteristics in the time domain. It has been found that these distinct features vary with the size of the droplets and its velocity. Experimental evidences show that the acoustic energy of the low frequency signal is related to the size of the drops that caused the acoustic noise and, hence, the energy of the drops generated acoustic signal is a measure of the drop size. An algorithm for detecting and computing the energy of these low frequency components leading to the estimation of DSD of rain is presented in this paper. This approach has been employed in developing a Disdrometer based automatic rain gauge, which is simple, reliable, and cost effective.

10:15

1aAOa7. Effects of scattering by air bubbles on performance of an underwater acoustic array. Richard Lee Culver and Mario F. Trujillo (Appl. Res. Lab. and Grad. Program in Acoust., Penn State Univ., PO Box 30, State College, PA 16804, rlc5@psu.edu)

Analytical and numerical calculations of the effects of nearby air bubbles on performance of an underwater acoustic array are presented. Array performance is characterized by the array gain. Two effects of air bubbles on array gain are considered: (1) attenuation of the direct path signal of interest and (2) additive, correlated interference due to scattering by nearby bubbles. The incident acoustic field is taken to be locally planar. Bubble scattering contributions to the interference field are calculated using the single scattering approximation of Ishimaru (1977, Chap. 6) assuming contributions by resonant bubbles only. Comparison is made between the acoustic effects of homogeneous and inhomogeneous (clumpy) distributions of bubbles. One feature of this work is that bubble distribution can be generated using computational fluid dynamics calculations for a fluid field and momentum transport equations for the bubbles. [Work sponsored by Office of Naval Research, Code 321 Undersea Signal Processing.]

Session 1aAOb

Acoustical Oceanography, Physical Acoustics, and Underwater Acoustics: Hank Medwin Memorial Session I

D. Vance Holliday, Chair

BAE Systems, 4669 Murphy Canyon Rd., San Diego, CA 92123-4333

Chair's Introduction—10:50

Invited Papers

10:55

1aAOb1. Wedge diffractions: Biot-Tolstoy normal coordinate solution and Medwin's lab experiments. Clarence S. Clay (Geology and Geophys., Univ. of Wisconsin, Madison, WI 53706)

In the 1960s, the Hudson Labs of Columbia University had coffee hours. The coffee gang included Ivan Tolstoy, Antares Parvulescu, Bob Frosch, Mel Hinich, and Peter Rona. We tested our crazy ideas, scribbled on the blackboard, and listened. Wave propagation in a messy ocean was the problem. We were well beyond ideal layered waveguides with plane interfaces. Tolstoy did normal coordinates and modes [Tolstoy and Clay, Ocean Acoustics (1966, 1986)-AO]. One summer, Hank Medwin joined us and conversations led to research on sound scatter from rough surfaces. Helmholtz-Kirchhoff (H-K) computations were accurate for sound scattered in specular directions, AO Ch. 6. Diffractions are different. The AO (1986) includes Biot-Tolstoy diffraction of impulses by rigid wedges (B-T). Interior wedge angles= ?/n have no diffracted waves. Clay and Medwin's Acoustical Oceanography (1977) detailed H-K methods and ignored B-T. Pushing hard, was H-K accurate for all diffraction waves? In 1978–82, Bremhorst, Jebsen, and Hank did wedge experiments and proved B-T is exact and H-K is limited. See Figs. 12.3, 2–3.3 in Medwin and Clay, "Fundamentals of Acoustical Oceanography" (1998). Chapters 11 and 12 give details. Use normal coordinate B-T for wedges.

11:15

1aAOb2. Blending underwater acoustics and acoustical oceanography. Timothy Duda (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Acoustical oceanography, which is the use of sound to probe ocean processes, had been in practice for many years before Hank Medwin spearheaded the creation of the ASA Acoustical Oceanography Technical Committee (AOTC). It has blossomed since that time, aided by interactions fostered by AOTC activities. The differences between acoustical oceanography and underwater acoustics are subtle and may not exist rigorously, particularly when propagation and scattering processes are exploited. However, the different foci of the two groups and the wealth of topics to be explored and researched each support the notion of two ASA groups devoted to ocean acoustics. In a similar fashion, AOTC interactions with many other committees foster progress. A review will be given of how underwater sound generation, propagation, and scattering knowledge, fundamental to UW, have been turned around to investigate the ocean, with a focus on work since the creation of the AOTC. Current trends in expeditionary oceanography and persistent monitoring will be covered, as well as the potential role of acoustics in future ocean observatory systems.

11:35

1aAOb3. Hank Medwin's early contributions to the use of acoustics in understanding physical processes in the ocean. Oscar B. Wilson (Phys. Dept., Naval Postgrad. School, Monterey, CA 93943) and Thomas G. Muir (Univ. of Mississippi, University, MS 38677)

Two old friends and colleagues of Hank Medwin will present a brief review of Hank's early acoustics research, during the period, 1960s–1980s, and its applications to processes in the ocean. Emphasis will be on work in the laboratory aimed at obtaining an understanding of the physics of some acoustical phenomena in the ocean; in particular, to diffraction from a sea mount, scattering from and transmission through a rough surface, measurement of bubble distribution in the ocean and its effects on sound propagation, as well as noise from rain drops impinging on the ocean surface.

Session 1pAA

Architectural Acoustics, Noise, Speech Communication, and ASA Committee on Standards: Advancements in Speech Privacy

Kenneth W. Good, Jr., Chair
Armstrong World Industries Inc., 2500 Columbia Ave., Lancaster, PA 17603

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. Recent Leadership in Energy and Environmental Design (LEED) and American Institute of Architects decisions about speech privacy: What do they mean? David Sykes (23 Buckingham St., Cambridge, MA 02138) William Cavanaugh, Gregory Tocci, and Andrew Carballeira (Cavanaugh-Tocci Assoc., Sudbury, MA 01773)

Fifty years of research and development in speech privacy that began in 1955 produced total demand for services and solutions by 2005 of less than $100 million a year. But suddenly this is changing due to the convergence of several forces that include: a surge of security and privacy laws over the past decade; growing concern about the noise epidemic in healthcare; and corporate concerns about privacy in business decision-making such as the SEC suit against Hewlett Packard. Result: recent decisions by the American Institute of Architects, LEED/U.S. Green Building Council, and the Green Guide for Healthcare that, for the first time, recognize speech privacy as an objective, measurable phenomenon. These decisions will produce demand for professional advice and counsel. Is the 8000-member acoustics profession prepared to meet a sudden increase in demand for services related to privacy measurement, mitigation, and certification? What needs to happen to ensure the profession continues to play a role in implementing and enforcing this basic but only recently recognized civil right?

1:25

1pAA2. Open plan privacy index measurement source speaker evaluation. Kenneth Good and Kenneth Roy (Armstrong Innovation Ctr., 2500 Columbia Ave., Lancaster, PA 17552)

Open Plan Privacy Index (PI) results have shown to have variability based on source speakers used for the field qualification measurements. Earlier studies have suggested that acoustic energy projected from the back of the source speaker may have an impact on this variability. The concern is that the ASTM E 1179 source speaker qualification required for ASTM E 1130 (Standard Test Method for Objective Measurement of Speech Privacy in Open Offices Using Articulation Index) is vague regarding the energy from the rear and sides of the source. Therefore, it is conceivable that different source speakers that qualify per E 1179 will produce different PI results due to differences in energy output from the back of the speaker, and the reflectivity of the subject environment. This study will discuss measurements using a variety of source speakers over a variety of open offices environments and how the speaker directivity differences impacted results.

1:45

1pAA3. The N number, a speech privacy metric for rating walls, revisited. Richard D. Godfrey, Harry Alter, and Clark Berdan III (Owens Corning, 2790 Columbus Rd., Granville, OH 43023)

In 1962, Cavanaugh, *et al.*, published the results of research which lead to a single number rating system for speech privacy of walls referred to as the N number ["Speech Privacy in Buildings," J. Acoust. Soc. Am. **34**, 475–492 (1962)]. The N number is derived from measured one-third octave band transmission loss data weighted using factors signifying the relative contribution to the articulation index. When the N number is summed with factors for source room floor area, source room speech use, measured adjacent room background noise rating, and the privacy requirement, the N number correlated subjective reactions observed in case studies better than average wall transmission loss alone. Owens Corning, the sponsor of this research, published the N number along with sound transmission class (STC) for several years. With the emergence of regulation of speech privacy in health care facilities, the Health Insurance Portability and Accountability Act (HIPAA), it seemed reasonable to revisit the N number. This paper will review the N number methodology and compare some building partition STCs and N numbers, with special attention on the impact of cavity insulation.

2:05

1pAA4. New techniques for measuring speech privacy and efficiency of sound masking systems. Peter Mapp (Peter Mapp & Assoc., Prospect House, 101 London Rd, Copford, Colchester, Essex C06 1LG, UK, Petermapp@btinternet.com)

Traditionally, speech privacy has been measured by means of the Articulation Index (AI), transposed to rate privacy rather than intelligibility (PI=1-AI). However, this is an indirect and slightly cumbersome method that usually requires a spreadsheet calculation to yield the Privacy Index rating. Furthermore, the traditional AI approach does not take into account room reverberation effects (though the very much more complex SII does). The paper discusses the potential use of STI and STIPa as direct measures of speech

privacy and shows that, potentially, they may be able to offer a superior approach. The benefits and limitations of the methods are highlighted together with the results from a number of case studies. It is concluded that while the method has potential merit, a number of the limiting factors require further research. It particular, it is shown that the index scale at lower levels of intelligibility may need to be revised.

2:25–2:35   Break

2:35

1pAA5. Practical design and assessment of architectural speech privacy for closed rooms.   Bradford N. Gover and John S. Bradley   (Inst. for Res. in Construction, Natl. Res. Council Canada, 1200 Montreal Rd., Ottawa, ON K1A 0R6, Canada, brad.gover@nrc-cnrc.gc.ca)

Practical procedures have been developed for predicting levels of speech privacy associated with closed meeting rooms. The procedures can be used to specify construction designs for rooms intended to provide desired levels of speech privacy. They can also be used for measurement, assessment, and rating of existing rooms. The new approach relates the predicted or measured sound insulation provided by the wall construction directly to the probability that speech will be audible or intelligible to bystanders outside a room. The audibility or intelligibility of speech depends on the relative levels of speech and background noise at the listeners position and can therefore be determined from the speech level inside the room, the sound insulation, and the noise level at the listening position. However, speech and noise levels vary from moment to moment. The measured statistics of these variations can be used to predict the probability that speech will be audible or intelligible outside the room. Wall constructions having higher sound insulation will result in lower probabilities of speech being overheard, and therefore correspond to higher degrees of speech privacy.

2:55

1pAA6. Measurements of speech privacy in healthcare.   Kenneth P. Roy and Kenneth W. Good, Jr.   (Armstrong World Industries, 2500 Columbia Ave, Lancaster, PA 17604)

Over the last two years a series of speech privacy measurements were conducted in two healthcare facilities including a hospital, a clinic, and a medical office building. Although an ASTM field measurement method (E-1130) currently exists for open plan spaces that can be applied to healthcare applications, such a method is not currently specified for closed plan spaces. At this time, our best option is to apply the process designated in ASTM E-1130 with the additional provisions of E-336 for room-to-room NR measurement. Test results will be presented from these evaluations, and these will be discussed relative to the research conducted at the National Research Council of Canada which has taken a statistical approach to resolving one of the primary issues—in-room speech levels.

3:15

1pAA7. The open office conundrum: Post-occupancy assessment of speech privacy in offices.   Kevin Powell   (US General Services Administration, 555 Battery St., Rm. 518, San Francisco, CA 94111, kevin.powell@gsa.gov), Charles Salter, and Randy Waldeck   (Charles M. Salter Assoc., Inc., San Francisco, CA 94104)

Post-occupancy evaluation is used to assess whether a completed building is delivering to its occupants and owners the benefits which were set as goals in the design process, and utilizing this information to improve future projects. As part of its workplace research program, the U.S. General Services Administration (GSA) has embarked upon an ambitious program where both objective measures of physical conditions (e.g., Articulation Index) and occupant satisfaction are assessed. Comprehensive evaluation of eleven GSA workplaces has shown that both before and after renovation, employees consistently rank acoustical privacy high in importance, but low in performance. This presentation will address three questions. First, what have we learned about the objective acoustical performance of contemporary open plan workplaces? Second, what have we learned about occupants subjective perception of office acoustics? Third, what guidance for new projects can GSA provide designers to improve occupant satisfaction with office acoustics, within the context of contemporary workstyles and current business objectives for the workplace, based on these post-occupancy evaluations?

3:35

1pAA8. A case study in measuring speech privacy in a flexible closed/open space.   Kenneth Good   (Acoust. Privacy Enterprises, LLC, P.O. Box 25, Mount Joy, PA 17552)

When we think about speech privacy we often think about the typical open plan office cubical or a small closed room such as a healthcare exam room. But an opportunity arose for an interesting case study with potential twists to the typical projects. This case study is for a flexible space within a church where a large common open room is used. This room is then divided off into several medium-sized Sunday school or meeting rooms in much the same concept as a hotel conference area—but with smaller rooms and low ceilings. The desire for privacy is to both minimize distraction from the adjacent space and provide a high level of confidentiality for sensitive conversations. This case study will explore the unique challenges in meeting and qualifying privacy expectations for this space.

Contributed Paper

3:55

1pAA9. The open office privacy calculator: A value engineering tool for acoustic professionals. Jonathan Kemp, Mark Bell (Cambridge Sound Management, 27 Moulton St. Cambridge, MA 02138, jdkemp@cambridgesoundmanagement.com), and Thomas Horrall (Acentech, Inc. Cambridge MA 02138)

Open-plan spaces consistently challenge architects and acoustic consultants to achieve reasonable speech privacy and low inter-office distraction. This paper presents a computerized tool to optimize design decisions, maximizing speech privacy while minimizing design effort and implementation cost. Speech privacy can be estimated through real-time calculation of the Speech Privacy Index (SPI) at a grid of points in a given simulated open office plan. The Open Office Privacy Calculator graphically shows how much workers in that open office would be distracted by any single speaking person and the impact of office environmental parameters on the level of distraction. The tool provides a direct graphical link between design decisions, cost, and acoustic performance for partition height, ceiling height, ceiling type, and the addition of sound masking. Further, this tool introduces the Distraction Index (DI), a measure of how many adjacent office talkers that a given office worker can routinely be distracted by ($SPI = 70\%$). The combination SPI/DI Calculator allows acousticians and architects to make informed value judgments, choosing design elements which have the greatest impact on speech privacy. The Calculator has been validated with measured data and by comparison with other existing and accepted models for predicting SPI in office spaces.

MONDAY AFTERNOON, 4 JUNE 2007

SALON A, 1:30 TO 3:00 P.M.

Session 1pAB

Animal Bioacoustics: Echolocation Signals and Behavior

James A. Simmons, Chair
Brown Univ., Dept. of Neuroscience, Providence, RI 02912

Contributed Papers

1:30

1pAB1. Jamming avoidance reveals segregation of processing for detection and ranging in echolocating bats. James A. Simmons (Dept. of Neurosci., Brown Univ., Providence, RI 02912, james_simmons@brown.edu), Mary Bates, and Sarah Stamper (Brown Univ., Providence, RI 02912)

Big brown bats emit frequency-modulated (FM) biosonar sounds that sweep from 50 to 23 kHz in the first harmonic. When searching for targets in the open or when detecting targets in wideband noise, bats lengthen their sounds and allow the FM sweep to tail off into a nearly constant-frequency (CF) segment, which boosts the energy in the sound over a narrow band of 24–28 kHz. When detecting a target in CF jamming sounds, bats shift their emitted frequencies up or down to move this terminal CF component from the interference when the CF frequency is 23–27 kHz. Other researchers have found a population of neurons in the bats inferior colliculus that is sharply tuned ($Q10dB = 30-40$) to frequencies of 24–30 kHz and may be used to detect echoes. However, these same frequencies also contribute incrementally to determining echo-delay accuracy for target ranging, and a parallel population of neurons has only moderately sharp tuning ($Q10dB < 5-10$). Sharper tuning in the cochlea might facilitate detection but could compromise time constants for reception, so these bats use neuronal inhibition in the inferior colliculus to create two parallel receivers at frequencies of 24–30 kHz. [Work supported by ONR and NIH.]

1:45

1pAB2. Strategy for vocalization taken by FM bats during group flight measured by a wireless microphone system. Hiroshi Riquimaroux, Yoshiki Osawa (Dept. of Knowledge Eng. & Computer Sci., Doshisha Univ., Kyotanabe, Kyoto 610-0321, Japan), and Yoshiaki Watanabe (Doshisha Univ., Kyotanabe, Kyoto 610-0321, Japan)

The purpose of this study was to examine characteristics of vocalization emitted by an FM bat, Japanese pipistrelle bats (Pipistrellus abramus), during their group flight. We used a wireless microphone system, Telemike, to record vocalization of five bats while they were flying in a chamber with a companion (a paired flight). Echolocation pulses and returning echoes were measured by the Telemike (0.6g; including a battery) placed at the bats head. Simultaneously, a pair of high-speed video cameras were used to reconstruct their flight trajectories. The Telemike recorded not only their own pulses and echoes but also those of the companion bat. Overlapping of their own echolocation signals with those of the companion bat were analyzed. For one out of two bats, overlapping was always much lower than chance level on average. On the other hand, overlapping was around chance level for the other one. Their strategy appeared that one bat reduced the overlapping while the other did not. Communication calls were also recorded between two individuals during flight, which were never recorded when the bats flew alone. [The research supported by a grant to RCAST at Doshisha Univ. from MEXT and by the Innovative Cluster Creation Project promoted by MEXT]

2:00

1pAB3. Rules-based front-end detector as a bootstrap method for model-based detection of microchiropteran calls. Mark Skowronski and Brock Fenton (Dept. of Biol., Univ. of Western Ontario, London, ON N6A 5B7, Canada, mskowro2@uwo.ca)

Rules-based methods for automated acoustical signal processing of bat calls have been developed throughout the history of bat acoustics research, stemming from techniques developed by expert acousticians for hand analysis of bat calls. Recently, a model-based paradigm [Skowronski and Harris, J. Acoust. Soc. Am. **119**(3), 1817–1833 (2006)], inspired by automatic speech recognition research, was introduced that improved the accuracy of detection and classification of echolocation calls by an order of magnitude over conventional techniques. However, the models require labeled data for training, and generating call end points by hand is time consuming and labor intensive. As an alternative, a rules-based front-end detector was developed to provide initial call end points for detection models. The combined system allows the models to be trained with unlabeled data, which greatly reduces the amount of time and effort needed to produce accurate detection models. With a sufficient amount of training data from across the spectrum of bat species, the model-based detector is superior to the rules-based detector and also generalizes to calls from species not present in the training data.

2:15

1pAB4. The acoustic function of flaps and furrows in the noseleaves of horseshoe bats. Rolf Müller, Qiao Zhuang, and Fuxun Wang (School of Phys. & Microelectronics, Shandong Univ., 5 Hongjia Lou, 250100 Jinan, China)

The nostrils of horseshoe bats are surrounded by elaborate noseleaves which consist of three parts: anterior leaf, sella, and lancet. Of these parts, shape and location of the anterior leaf, which forms a baffle around the nostrils, are most suggestive of an acoustic function. While such an argument is not as readily made for them, the other two parts show more conspicuous structural features than the anterior leaf. In particular, the lancet is furrowed by deep horizontal grooves in most species and flaps can be attached to the bottom of the lancet in some species. The results of a numerical analysis presented here indicate that the lancet furrows and the sella flaps both have effects on the animals' radiation pattern. The acoustic mechanisms behind these effects differ; however, Lancet furrows act as half-open resonance cavities whereas the sella flaps have a more complicated impact on the wavefield phase and amplitude. Despite the differences in the acoustic mechanisms, there are two important commonalities in the effects of these structural features: Both act as devices for beam dispersal widening the distribution of sound energy over angle and the effects of both are limited to the FM-component of the biosonar pulse.

2:30

1pAB5. Sound localization and its relation to vision in large and small New-World bats. Gimseong Koay, Henry E. Heffner, and Rickye S. Heffner (Dept. of Psych. MS 948, Univ. of Toledo, 2801 W. Bancroft St., Toledo, OH 43606, Rickye.Heffner@utoledo.edu)

Passive sound-localization acuity (minimum audible angle) for brief noise bursts was determined behaviorally for two species of New-World bats (Phyllostomidae): *Phyllostomus hastatus*, a large bat that eats fruit and preys on other vertebrates, and *Carollia perspicillata*, a small species that eats fruit and nectar. Both use echolocation calls of very low intensity

for orientation and obstacle avoidance. The mean minimum audible angle for two *P. hastatus* was 9 deg, and that for two *C. perspicillata* was 14.8 deg. This places their passive sound-localization acuity near the mean for mammals. Sound localization varies widely among mammals and the best predictor of a species acuity remains the width of the field of best vision ($r = .89$, $p < .0001$). Neither of these bats nor the other three bat species tested so far deviates from the relationship between sound-localization and vision, suggesting that despite their specialization for echolocation, the use of hearing to direct the eyes to the source of a sound still serves as the major source of evolutionary selective pressure for passive sound localization.

2:45

1pAB6. Modeling and analysis of the backscatter of dolphin echolocation clicks from elastic targets. Mark W. Muller, John S. Allen III (Dept. of Mech. Eng., Univ. of Hawaii, 2540 Dole St., Honolulu, HI 96822), Whitlow W. L. Au, and Paul E. Nachtigall (Marine Mammal Res. Program, Kaneohe, HI 96744)

The clicks from an Atlantic bottlenose dolphin were obtained in a controlled experiment at the Marine Mammal Research Program (Kaneohe Bay, HI). A previously described phantom echo experiment using solid elastic spherical targets provides the foundation for an investigation of the specific scattering components used in a dolphin's target identification and classification methods. A model of the scattering interaction of incident dolphin echolocation clicks and elastic targets is outlined. The frequency spectra from the echolocation clicks are combined with the resonance-scattering representations of the scattered pressure fields to yield filter-type integrals that can be viewed as filtering the form function of the scatterer through the spectral window of the incident echolocation clicks. This analysis is carried out in the time and frequency domains and also represented in time-frequency distributions (Wigner-Ville, RID). An advantage of this analysis is that it can be carried out for any arbitrary echolocation click incident on targets of any general shape and composition. The results are compared and discussed with respect to experimentally obtained back-scattered signals.

MONDAY AFTERNOON, 4 JUNE 2007

CANYON ROOM B/C, 1:30 TO 5:20 P.M.

Session 1pAO

Acoustical Oceanography, Physical Acoustics, and Underwater Acoustics: Hank Medwin Memorial Session II

David M. Farmer, Chair
Univ. of Rhode Island, Graduate School of Oceanography, South Ferry Rd., Narragansett, RI 02882

Chair's Introduction—1:30

Invited Papers

1:35

1pAO1. Experiments on the origin of breaking wave noise. Grant B. Deane (Scripps Inst. of Oceanogr., Code 0238, UCSD, La Jolla, CA 92093-0238)

The origin of wave noise has been of interest since the works of Wenz and Knudsen in the first half of the last century and remains an active area of research. Medwin made seminal contributions to this topic. His measurements of the sound generated by gently spilling breakers, along with those of Banner and Cato, settled the question of the origin of the noise: bubbles. An issue that remains unresolved is the nature of the excitation mechanism at bubble formation leading to acoustic emission. This will be discussed along with some recent laboratory measurements of fragmenting bubbles. [Work supported by ONR.]

1:55

1pAO2. The effect of a rough sea surface on acoustic normal modes. James H. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882) and James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Acoustical oceanography as defined by Medwin is concerned with development and use of acoustical techniques to measure and understand parameters and processes of the sea. The book of the same name, by Clay and Medwin, was used in a graduate course at MIT for years and educated a great many young acousticians about sound in the ocean including the effects of its surface. Stressed by the turbulent winds above it, the ocean surface can be tweaked to tiny patches of cats paws or whipped into a frenzy of raging, mountainous seas. [Clay and Medwin, "Acoustical Oceanography" (Wiley, New York, 1977) p. 16]. The text developed the fundamentals of ocean acoustic propagation including the normal mode solution in a waveguide. With the background provided by this text, we developed a technique for estimating the sea surface frequency-direction spectra using the perturbations in modal travel time. The forward problem of computing the acoustic travel time perturbation spectra given the surface wave spectrum was solved to first order. The recent Shallow Water 06 experiment experienced two tropical storms (Ernesto and Florence) and mode travel time perturbations were measured between a number of sources and receivers and are compared to theory. [Work supported by ONR.]

2:15

1pAO3. Medwin's maxim for finding good research questions: Listen to your students. Or a freshman's approach to shallow water acoustics. John Colosi (Dept. of Oceanogr., Naval Postgrad. School, Monterey, CA 93943)

Over the years, Hank Medwin worked with numerous students and postdoctoral scientists, making fundamental discoveries in almost every branch of ocean acoustics. Hank's motivations were not so altruistic, but were based on a deep appreciation for the fact that young scientists would ask simple questions that the more seasoned scientist often did not. As a young faculty member of the Naval Postgraduate School, Hank wasted no time in passing this information onto me, the result of which is the freshman's approach to shallow water acoustics discussed in this talk.

2:35

1pAO4. Hypotheses on the exploitation of bubble acoustics by cetaceans. Timothy Leighton, Daniel Finfer, and Paul White (Inst. of Sound and Vib. Res., Univ. of Southampton, Highfield, Southampton SO17 1BJ, UK)

Man-made sonar does not operate well in bubbly water, and yet cetaceans not only function effectively in shallow coastal waters, but also at times generate large bubble fields to assist with catching prey. This paper outlines the challenges faced by cetaceans in using acoustics in such environments, and proposes acoustical techniques which would work. The validity of such proposed acoustical solutions is explored through theory, simulation, and experimentation. The scenarios in question relate to the circular and spiral bubble nets generated to trap prey by humpback whales, and solutions to difficulties associated with echolocation by dolphins and porpoises in bubbly water. Whether the solutions are exploited by cetaceans is uncertain, but their efficacy in test tanks and implications for man-made sonar are demonstrated.

2:55

1pAO5. Sorting out the racket: Acoustic footprints for ocean climate. Jeffrey A. Nystuen (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, nystuen@apl.washington.edu)

Ocean ambient sound is a component of the physical marine environment that contains quantifiable information about physical processes, and patterns of biological and anthropogenic activity. Sorting out the different sound sources is a necessary component for utilizing ambient sound to monitor the environment. Hank Medwin played a fundamental role investigating the micro-physics of sound production by breaking waves and raindrops. These two sound sources, wind and rain, form the basis for distinctive acoustic footprints for different marine environments. During the past decade, passive aquatic listeners have been deployed from moorings in many different marine environments including deep ocean locations (SE Pacific, NE Pacific, Eastern and Western tropical Pacific), marginal seas (Bering and Ionian), and coastal/inland waterways (Cape Flattery, Puget Sound). Sound budgets developed for these different marine environments show distinctive patterns associated with different types of rainfall and storms present, and contain distinctive patterns of biological and anthropogenic activities. These sound budgets quantify acoustic ocean climate and can be used to monitor future changes in the marine environment. [Work supported by ONR Ocean Acoustics and NSF Physical Oceanography.]

Contributed Paper

3:15

1pAO6. Herman Medwin and forward scattering from rough surfaces near grazing incidence. Gerald L. D'Spain (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 93940-0701)

Herman Medwin, "The Professor," was a man of great energy who thrived on getting things done. One small indication of that were the athletic shoes he wore with his sports coat and tie when he attended Acoustical Society of America meetings; they allowed him to stay "on the run." I worked with him and Ocean Acoustics Associates for a 3 year period in the early 1980's. Our major research efforts then were devoted to laboratory-scale measurements of sound scattering from rough surfaces, in particular forward scattering near grazing incidence at low frequencies. A significant amount of the research was conducted by U.S. Navy officers in the physics department at the Naval Postgraduate School. The Professor had a gift of organizing the work into 6 month blocks appropriate for each student's master's thesis. A number of distinguished civilian scientists also passed through the laboratory. He was tremendously proud of Fundamentals of Acoustical Oceanography and liked to say that the book, co-authored with his friend Clarence Clay, put him on the map. In the Acknowledgments in my Ph.D. thesis appears the sentence ["I want to specially thank The Professor, Herman Medwin, who helped put the fun back into science."]

Invited Paper

3:45

1pAO7. Herman Medwin as observed by an interested bystander. Wayne M. Wright (Appl. Res. Labs, Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Herman Medwin was a brilliant scientist, a complex individual, and a strong-willed advocate. Although his most significant scientific contributions were associated with acoustical oceanography, his professional creativity was not limited to this technical field. Hank's graduate work at UCLA involved acoustic streaming in gases; in his early professional years he reported substantial work with intense sound waves in gases; he was an accomplished violinist; and his final published work involved auditorium acoustics. His contributions to the Acoustical Society of America were particularly noteworthy, from his initial service as a technical committee and Executive Council member through its presidency. His well-deserved recognition by ASA included award of the Silver Medal in Acoustical Oceanography and the Gold Medal. He spearheaded establishment of the Medwin Prize in Acoustical Oceanography, which is awarded annually by ASA to provide recognition to mid-career researchers in the field. Personal observations are given on some of these topics by an outsider (from AO), who was privileged to spend a year assisting with Hank's research, and who closely followed his experiences with ASA through the formalities associated with creation of the prize given in his name.

Contributed Papers

4:05

1pAO8. High-frequency broadband (150–600 kHz) acoustic backscattering from oceanic microstructure and zooplankton in the presence of internal solitary waves. Andone C. Lavery and Dezhang Chu (Woods Hole Oceanogr. Inst., Dept. of Appl. Ocean Phys. and Eng., MS 11, Woods Hole, MA 02543)

As one of the founders of acoustical oceanography, Hank Medwin has directly or indirectly influenced the careers of many scientists. Throughout his illustrious and prolific career, Hank Medwin was passionate about using and developing acoustic techniques to study the ocean and, in particular, to study the physical processes of the ocean interior, a passion shared by the authors. In this spirit, the authors present highlights of an experiment in which a high-frequency broadband (150–600 kHz) acoustic backscattering system was used to investigate the scattering of sound from oceanic microstructure and zooplankton throughout the evolution, from generation to dissipation, of internal solitary waves. The broadband spectrum of the scattered returns is used to differentiate between regions in which the scattering is dominated by zooplankton versus microstructure. Interpretation of the measured scattering relies heavily on a fundamental understanding of scattering processes as well as on direct measurements of microstructure and biology.

4:20

1pAO9. High frequency signal amplitude fluctuations in shallow water. Nicholas P. Chotiros, Marcia J. Isakson, James N. Piper (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029), and Mario Zampolli (NATO Undersea Res. Ctr., La Spezia, Italy)

Amplitude fluctuations were measured in a high-frequency (5–50 kHz) experiment over short ranges in 10 meters of water on the north shore of Isola dElba, Italy. The fluctuations, as a function of range and frequency, are compared with estimates using the empirical refractive index wave-number spectrum put forward by Medwin [J. Acoust. Soc. Am. **56**, 4 (1974)] for the upper ocean region. The spectrum is governed by the distance from the nearest boundary and the standard deviation of the refractive index. The results illustrate the ubiquity of refractive index fluctuations and the elegance of Medwin's model. [Work supported by Office of Naval Research, Ocean Acoustics.]

4:35

1pAO10. Interaction of an inhomogeneous acoustic plane wave with a fluid sphere: Analytical solution. Kenneth G. Foote (Woods Hole Oceanograph. Inst., 98 Water St., Woods Hole, MA 02543)

The problem of scattering of a plane acoustic wave by a fluid sphere was addressed by V. C. Anderson [J. Acoust. Soc. Am. **22**, 426 (1950)]. This wave was assumed to be homogeneous, without amplitude variation over the wave front. The problem of scattering of a plane wave partially insonifying a rigid sphere was addressed by G. C. Gaunaurd [IEEE J. Oceanic Eng. **10**, 213 (1985)], but in the Kirchhoff approximation. Here, the problem of scattering of an inhomogeneous plane acoustic wave by a fluid sphere is addressed. A general analytical solution is given. A particular solution is described for a monochromatic plane wave whose amplitude on the surface of a sphere of radius r=a, with center at the origin, is $\exp[ika \cos \theta - 2\alpha a \cos^2(\theta/2)]$, where k is the wavenumber, θ is measured relative to the direction of propagation, and α is the absorption coefficient. Both the scattered and internal fields are evaluated for a range of parameter values of k, α, r, and θ, and compared against the Anderson solution for a homogeneous plane wave. It is believed that H. Medwin would have appreciated the unexpressed direction of this work: modeling of sonar interactions with marine animals.

4:50

1pAO11. A new multiple angle scattering apparatus for fish and zooplankton studies. Paul L. D. Roberts and Jules S. Jaffe (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

Recent computer studies have indicated that a substantial improvement in classifying zooplankton and sizing fish bladders could be obtained by measuring sound scatter at a multiplicity of angles. In order to investigate the application of this principle to sound scattering from live animals, we have developed a new multiple angle scattering apparatus. The system consists of a linear array of high-frequency, broadband transducers that are aligned so that all beams intersect at roughly the same location. Complementing the acoustic sensors, the system employs two low cost cameras focused on the volume insonified by the array. The camera images are used to estimate the size and orientation of the scatterer during insonification. Preliminary results show dramatic differences in scatter recorded at each angle. These data demonstrate that recording multiple angle scatter can yield novel information about the animal being insonified, which may be used to improve estimates of its size, orientation, and taxa.

5:05

1pAO12. Numerical modeling of groundfish target strength compared to *in situ* data. Sairajan Sarangapani, Jennifer L. Miksis-Olds (School for Marine Sci. and Technol., Univ. of Massachusetts, 838 S. Rodney French Blvd., New Bedford, MA 02744, ssarangapani@umassd.edu), and D. Benjamin Reeder (Naval Postgrad. School, Monterey, CA 93943)

Acoustic scattering models of zooplankton and fish are an essential component of the process of converting *in situ* target strength of marine organisms into meaningful biological parameters such as size and numerical density. This study compared the numerical model of axisymmetrically shaped groundfish target strength (e.g., cod, dog fish, haddock) with measurements obtained *in situ*. Experimental target strength data were ob-

tained using a bottom-mounted 200-kHz acoustical platform, fitted with four laser beams and an optical recording system to correlate species, length, and incidence angle to the measured target strength. Based on the length of the groundfish obtained from the data *in situ*, target strength modeling estimates were computed using the FMM (Fourier matching method), assuming an axisymmetrically shaped swim bladder with soft boundary conditions. The scattering coefficients for the truncated modal series solutions were predicted in quad precision, which added stability to the results. Data-model comparisons of target strength at 200 kHz are presented as a function of incidence angle for a number of groundfish of different lengths and species. [Work supported by NOAA.]

Session 1pNS

Noise: Engineering Noise Control for the Mining Industry

Michael Hankard, Cochair

Hankard Environmental Inc., 211 East Verona Avenue, Verona, WI 53593

John P. Seiler, Cochair

Mine Safety and Health Administration, Cochran Mill Rd., P. O. Box 18233, Pittsburgh, PA 15236

Chair's Introduction—2:00

Invited Papers

2:05

1pNS1. Mining in the U.S. Rocky Mountain West (circa 2007) and its environmental noise implications. Michael Hankard and Jeff Cerjan (Hankard Environ., Inc., 306 Rose Finch Cr., Littleton, CO 80129)

There's gold in them hills, and some oil, gas, and coal, too. Commodity prices are up, the desire to develop domestic energy supplies is strong, and the West has some (relatively) green fuels to be had. The relevance of this to acoustics is that it takes a lot of big, noisy equipment to get these resources to market. This, combined with the construction of homes advancing further into prime mining lands, tends to create noise conflicts between landowners and mineral rights owners. This study identifies the locations of existing mining operations in the West, and looks at which sectors are on the rise. The major noise producing activities associated with various mining activities are described, including earth moving, processing, transportation, and blast noise and vibration. A review of noise propagation models, applicable environmental noise regulations, and mitigation strategies is also provided. This report is intended to be a useful reference for those interested in gaining an understanding of the location and nature of current and projected mining operations in the U.S. West, the noise generated by these operations, and how noise conflicts between mining and homeowners are being addressed.

2:25

1pNS2. Operator noise exposure reduction using coated conveyor chain flight bars. Adam K. Smith, Peter G. Kovalchik, and Lynn A. Alcorn (Natl. Inst. for Occupational Safety and Health, 626 Cochrans Mill Rd., Pittsburgh, PA 15236, asmith9@cdc.gov)

Continuous miners are one of the fundamental machines used to cut, gather, and remove coal and noncoal minerals from underground room and pillar mining facilities. These machines are among mining equipment that generates the highest sound levels in an underground environment. The Mine Safety and Health Administration has determined that continuous mining machines rank first among all equipment in underground coal mining whose operators exceed 100% noise dosage. The conveying system is one of the principal noise sources on continuous mining machines, due to metal on metal impacts that occur between chain flights and the conveyor deck. A highly durable polyurethane coating has been developed by NIOSH for the chain flights to decrease noise generated by these impacts. A continuous mining machine retro-fitted with coated flight bars have achieved overall sound level reductions of 5–7 dB(A) in a laboratory setting (Kovalchik *et al.*, 2004). This research describes the effectiveness of this engineering noise control in reducing the noise exposure of continuous mining machine operators in an underground environment. Results show that continuous mining machine operators receive a 3 dB(A) time-weighted average reduction over an 8-h period.

2:45

1pNS3. Noise controls for drills in the dimension stone industry. Anthony S. Argirakis (U.S. Dept. of Labor, Mine Safety & Health Admin., Pittsburgh Safety & Health Technol. Ctr., P.O. Box 18233, Cochrans Mill Rd. Pittsburgh, PA 15236)

Drills in the dimension stone industry have proven to be a problem when it comes to overexposures to noise. These drills include, but are not limited to slot drills, down-hole drills, and horizontal lift-hole drills. Although drills with acoustically treated operator cabs exist for this industry, most drills encountered do not have them. Considerable effort has been made by both the Mine Safety and Health Administration (MSHA) and industry to achieve compliance with the noise standards stated in Code of Federal Regulations (CFR) Title 30, part 62 for the drills without operator cabs. Several individual noise controls are available for these drills. They include exhaust mufflers, operator control barriers, hammer covers, face shields, and three-sided enclosures. A joint noise study was

conducted by MSHA and a cooperative mining company to determine the effectiveness of these individual noise controls alone and in various combinations. Previously, these noise controls were studied separately under varying conditions and some found to provide significant noise reductions. The study was conducted to determine their effectiveness under the same conditions and to quantify not only their reduction in sound level, but also overall noise dose.

3:05

1pNS4. Drill steel research to reduce roof bolting machine operator noise exposure. Jeffrey Shawn Peterson (NIOSH, 626 Cochran's Mill Rd., P.O. Box 18070, Pittsburgh, PA 15236)

NIOSH has devoted significant resources to reducing Noise Induced Hearing Loss (NIHL) among the Nation's miners. As part of this effort, NIOSH's Pittsburgh Research Laboratory (PRL) employs a large reverberation chamber for sound power level measurements and a hemi-anechoic chamber and beamforming system for noise source identification. One of several research projects at PRL is investigating noise controls to reduce the noise exposure of roof bolting machine operators. Research findings to date suggest that drill steel vibration is a significant source of the noise emissions generated during the bolting/drilling cycle and a major contributor to the operators' noise exposure. In this presentation, the laboratory and field test methodologies are explained as well as a summary of the data analysis procedures and testing results. Proper implementation of currently available noise controls can reduce sound levels at the operators' position from 3 to 7 dB(A) under laboratory conditions. Conclusions drawn from these results and the steps taken to generate effective engineering noise controls for reducing roof bolting machines operators' noise exposures are also covered.

3:25

1pNS5. Composite saw blade technology. John P. Homer and Leonard C. Marraccini (U.S. Dept. of Labor–Mine Safety and Health Administration (MSHA), 626 Cochrans Mill Rd., Pittsburgh, PA 15236)

Stone saw blades are a significant problem concerning noise in the mining industry. Composite blade core technology provides the means for reducing the amount of noise emission from cutting processes. Copper/steel laminate construction has demonstrated the ability to significantly reduce noise emission by 9.2–12.7 dB. This type of noise control technology has been investigated by MSHA to be published as a proven noise control technology. The purpose of this investigation is to classify and quantify the means by which noise reduction is achieved by a composite blade core. The alternate construction of the blade core serves to increase the mechanical hysteresis of the structure. This lowers the vibratory response of the blade and thereby reduces sound emission. Many methods of evaluation are employed in this analysis. The standard method is a comparison of sound pressure level and full-shift noise dose for a steel and composite blade. Further, vibratory decay measurements were taken to predict sound reduction. Modal testing was also performed to reveal potential resonant operating conditions and to observe the effects of mounting configuration on system response. The results of these analyses provide substantial proof that composite saw blade technology provides significant noise reduction for stone cutting applications.

Contributed Paper

3:45

1pNS6. Control of 24-h blasting noise during construction of light rail tunnels in Portland, OR. Kerrie G. Standlee (Daly-Standlee & Assoc., Inc., 4900 SW Griffith Dr., Ste. 216, Beaverton, OR 97005)

In 1993, the Tri-County Metropolitan Transportation District of Oregon (Tri-Met) began construction of the Westside Light Rail project in Portland, OR. The project included the construction of a 3-mile-long twin tunnel through the hillside west of downtown. The original plan called for the tunnels to be excavated from the east using a tunnel boring machine that was supposed to be quick and quiet. However, soon after excavation began, the composition of rock along the tunnel path began to cause problems for the boring machine. Excavation fell behind schedule. Tri-Mets contractor decided that a second excavation team needed to move from the west end of the tunnels toward the boring machine to ensure the project would stay on schedule. Tri-Met requested permission to allow the contractor to conduct 24-h blasting in the vicinity of residences to help catch up in the excavation phase. Daly-Standlee & Associates, Inc. (DSA) assisted in defining noise criteria that could be used to ensure blasting noise would be acceptable at residences and, once blasting began, DSA assisted in developing mitigation measures that could ensure the criteria were met. This paper presents the experience gained by the firm during the tunnel boring project.

Session 1pSC

Speech Communication: Clear Speech and Auditory−Visual Processing (Poster Session)

Ann R. Bradlow, Chair

Northwestern Univ., Dept. of Linguistics, 2016 Sheridan Rd., Evanston, IL 60208

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-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

1pSC1. Global temporal characteristics of English clear and conversational speech. Rajka Smiljanic, Josh Viau, and Ann Bradlow (Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, rajka@northwestern.edu)

This study investigated the effect of hyperarticulated, intelligibility-enhancing clear speech on global temporal characteristics in paragraph-length utterances. The results of sentence-in-noise listening tests showed a consistent clear speech intelligibility gain across the utterances indicating that the talkers successfully maintained clear speech articulatory modifications throughout the paragraphs. We then explored temporal properties of the speech signal in terms of the percentage and variability of vocalic and consonantal intervals (following Ramus *et al.*, 1999). Consonantal and vocalic intervals were lengthened equally in clear speech: %V remained stable across speaking styles. Moreover, coefficients of variation for both consonantal and vocalic intervals remained stable across clear and conversational speech, suggesting that global temporal properties remain rather stable in the two speaking styles. There was also evidence of an increase in the number of prosodic phrases and of vocalic and consonantal intervals for clear speech. Longer durations of consonantal segments in combination with insertion/strengthening of short consonantal segments that were dropped or coarticulated in conversational speech were reflected in an increased consonantal standard deviation in clear speech. Overall, these results suggest that increased intelligibility of clear speech may be attributed to prosodic structure enhancement (increased phrasing and enhanced segmentability) and stable global temporal properties.

1pSC2. Phonemic contrast enhancement in clear speech production: The case of Korean stop production. Kyoung-Ho Kang (Dept. of Linguist., Univ. of Oregon, 1290 Univ. of Oregon, Eugene, OR 97403)

This study explores the code enhancement aspect of clear speech production. Younger ($n=11$, 20–29 years old) and older ($n=11$, 40–60 years old) speakers of Korean produced the three series of Korean stops (aspirated, lenis, fortis) in three different speaking styles: conversational, citation-form, and clear speech. The production of Korean stops was analyzed for three acoustic properties differentiating the Korean stops, voice onset time (VOT), amplitude difference between the first and the second harmonics (H1-H2), and fundamental frequency (F0). For both of the younger and older groups, the VOT difference between the aspirated and lenis stops was enhanced in clear speech, in addition to enhanced difference between the aspirated/lenis and fortis stops for H1-H2. The two groups differed in that only the younger speakers showed a substantial increase in F0 for the production of the aspirated stops in clear speech. Also, there was a substantial degree of interspeaker variability in the size and direction of the acoustic adjustment in clear speech. The results of this study support Silva's [Phonology **23**, 287–308 (2006)] proposal that

younger and older Korean speakers may use different strategies in order to maintain a distinction between the three types of Korean stops.

1pSC3. Acoustic correlates of talker-reported clear strategies. Sarah Hargus Ferguson, Meredith A. Poore (Dept. of Speech-Lang.-Hearing: Sci. and Disord., Dole Ctr., 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045), and Rahul Shrivastav (Univ. of Florida, Gainesville, FL 32611)

When asked to speak clearly, talkers vary considerably in terms of the amount of intelligibility benefit achieved. The acoustic differences between clear and conversational speech also differ from talker to talker. The present study examined the extent to which clear speech acoustic modifications were consistent with talkers' descriptions of the strategies they used in the clear speech condition during the recording of a large multi-talker database [S. H. Ferguson, J. Acoust. Soc. Am. **116**, 2365–2373 (2004)]. Specifically, talkers who did and did not report placing extra stress on keywords in clear speech were compared on a variety of duration, pitch, and amplitude metrics. Results will be discussed in terms of talker differences in awareness of and control of specific aspects of speech production. Clinical implications will also be discussed.

1pSC4. Perception of clear English fricatives. Kazumi Maniwa (Fachbreich Sprachwissenschaft, Univ. Konstanz, F521A, 78457 Konstanz, Germany), Allard Jongman (Univ. of Kansas, Lawrence, KS 66044), and Travis Wade (Univ. of Stuttgart, 70174 Stuttgart, Germany)

Two experiments investigated whether and how clear speech production enhances intelligibility of English fricatives for normal-hearing listeners and listeners with simulated hearing impairment. Fricative intelligibility in clear and conversational speech was assessed using a database of 8800 VCV ([a]-fricative-[a]) stimuli produced by 20 speakers. Babble thresholds were measured for minimal pair distinctions for 14 normal-hearing listeners and 14 listeners with simulated sloping, recruiting impairment. Clear speech benefited both groups overall; however, for impaired listeners, the clear speech effect held only for sibilant pairs. In a previous acoustic study involving the same stimuli [Maniwa, Jongman, and Wade, J. Acoust. Soc. Am. **119**, 3301 (2006)], 14 parameters were measured: spectral peak location, the first four spectral moments, F2 onset transitions, spectral slopes below and above typical peak locations, pitch of adjacent vowels, overall rms amplitude, relative amplitude (frication amplitude relative to the vowel in specific frequency bands), harmonic-to-noise ratio, energy below 500 Hz, and fricative duration. Correlation analyses comparing acoustic and perceptual data indicated that a shift of energy concentration toward higher frequency regions and greater source

strength contributed to the clear speech effect for normal-hearing listeners, while listeners with simulated loss seemed to benefit mostly from cues involving lower frequency regions.

1pSC5. Clear speech effects for vowels produced by monolingual and bilingual talkers. Teresa DeMasi, Catherine L. Rogers, and Jean C. Krause (Dept. of Commun. Sci. and Dis., Univ. of South Florida, 4202 E. Fowler Ave. PCD1017, Tampa, FL 33620)

The present study investigates the hypothesis that bilinguals may produce a smaller intelligibility benefit than monolinguals when asked to speak clearly. Three groups of talkers were recorded: 13 monolingual native English speakers, 22 early Spanish-English bilinguals, with an age of onset of learning English (AOL) of 12 or earlier, and 14 later Spanish-English bilinguals, with an AOL of 15 or later. Talkers produced the target words "bead, bid, bayed, bed, bad" and "bod" in both clear and conversational speech styles. Two repetitions of each target word were mixed with noise and presented to monolingual English-speaking listeners in a six-alternative forced-choice task across two days of testing. Stimuli were also presented in quiet on two subsequent days of testing. In preliminary data from 13 listeners, the early bilinguals were slightly more intelligible in noise than the monolingual talkers, with both groups showing a similar degree of clear speech benefit. Later bilinguals were less intelligible overall and showed a much smaller clear speech benefit in noise. These data suggest that later bilinguals, but not early bilinguals, may experience a disadvantage when speaking in noise due to a reduced ability to improve intelligibility by speaking more clearly [Work supported by NIH].

1pSC6. An acoustic study of real and imagined foreigner-directed speech. Rebecca Scarborough, Olga Dmitrieva, Lauren Hall-Lew, Yuan Zhao (Stanford Univ., Linguist. Dept., Stanford, CA 94305-2150, rscar@stanford.edu), and Jason Brenier (Univ. of Colorado, Linguist., Boulder, CO 80309-0295)

Foreigner-directed speech is a much cited speech style, but its acoustic properties are surprisingly under-studied, and many existing studies evoke imagined interlocutors to elicit foreigner-directed speech. This study provides an acoustic comparison of foreigner-directed and native-directed speech in real and imaginary conditions. Ten native English speakers described the path between landmarks on a map to each of two confederate listeners (one native English speaker and one native Mandarin speaker) and to two imagined listeners (one described as a native English speaker and the other as a non-native speaker). Vowel duration, rate of speech, and vowel centralization were examined across native-foreigner and real-imagined conditions in 22 target words. Stressed vowels were longer and rate of speech was slower in foreigner-directed than in native-directed speech. Additionally, /ae/ was more peripheral in the vowel space in the foreigner-directed condition. These effects were consistent across real and imaginary conditions; however, vowels were longer, rate of speech was slower, and /ae/ was more peripheral in the imaginary than in the real condition. In other words, speakers made acoustic-phonetic adjustments in foreigner-directed speech that are consistent with those seen in listener-directed clear speech, and these accommodations were greater when the listener was imagined rather than real.

1pSC7. Early stages of audiovisual speech processing—a magnetoencephalography study. Ingo Hertrich, Hermann Ackermann (Dept. of General Neurology, Univ. of Tuebingen, Hoppe-Seyler-Str. 3, D-72076 Tuebingen, Germany, ingo.hertrich@uni-tuebingen.de), Klaus Mathiak (Univ. Hospital Aachen, D-52074 Aachen, Germany), and Werner Lutzenberger (MEG Ctr., Univ. of Tuebingen, D-72076 Tuebingen, Germany)

Speech is a multimodal process integrating auditory and visual information. It is still unsettled at which stage of processing visual and auditory data structures merge into a common percept. This study investigates the time domain of auditory M50 and M100 fields in response to audiovisual speech (whole-head MEG, 25 subjects). As in the original video recordings (/pa/ and /ta/), the visual articulatory cues preceded the acoustic signal (synthetic syllable) by ca. 150 ms. Visual nonspeech, acoustic nonspeech, visual-only, and acoustic-only stimuli served as control conditions. In accordance with previous studies, the M50 was consistently attenuated by visual motion cues, and a similar visual "negativation" effect could be observed even when no acoustic signal was played. As concerns the M100 field, visual nonspeech stimuli had an attenuating effect (only in the presence of an acoustic signal) whereas visual speech gave rise to hypoadditive M100 enhancement. Application of a six-dipole model showed a categorical-like visual /p/ versus /t/ effect. Since the source of this effect was located within insular cortex rather than the auditory system, we may assume that at the time of M100 the visual phonetic information has not yet been integrated into a common auditory/phonetic representation.

1pSC8. Frequency band-importance functions for auditory and auditory-visual sentence recognition. Ken W. Grant and Joshua G.W. Bentstein (Walter Reed Army Medical Ctr., Army Audiol. and Speech Ctr., Washington, DC 20307-5001)

For consonant recognition, the frequency band-importance function (FBIF) for audio-alone (A) and auditory-visual (AV) speech recognition are significantly different. The relative importance of very low speech frequencies is greater for AV than for A conditions, whether individual frequency bands with different center frequencies are presented in isolation or simultaneously. However, sentences filtered into different frequency bands of approximately equal intelligibility yield no difference between the A and AV FBIFs [Grant and Braida, J. Acoust. Soc. Am. **89**, 2952–2960 (1991)]. A correlation method designed to estimate frequency weighting functions for spectral profile analysis, but applied to speech, yields very different FBIFs under broadband versus isolated-band conditions [Turner *et al.*, J. Acoust. Soc. Am. **104**, 1580–1585 (1998)]. The present study sought to reexamine the A and AV FBIFs for sentences using broadband speech. IEEE sentences were filtered into three simultaneously presented non-overlapping frequency bands adjusted for equal A intelligibility. Stimuli were presented in noise filtered into the same bands as the speech, with the signal-to-noise ratio (SNR) selected randomly and independently for each band. The FBIF was estimated by calculating a correlation coefficient between performance and the SNR in each band. Implications for hearing-aid design and fitting will be discussed.

1pSC9. Temporal characteristization of auditory-visual coupling in speech. Adriano V. Barbosa, Hani C. Yehia, and Eric Vatikiotis-Bateson (E270, 1866 Main Mall, Vancouver, BC V6T 1Z1, Canada, evb@interchange.ubc.ca)

This work examines the coupling between the acoustic and visual components of speech as it evolves through time. Previous work has shown a consistent correspondence between face motion and spectral acoustics, and between fundamental frequency (F0) and rigid body motion of the head [Yehia *et al.* (2002), JPHON, **30**, 555–568]. Although these correspondences have been estimated both for sentences and for running speech, the analyses have not taken into account the temporal structure of speech. As a result, the role of temporal organization in multi-modal speech cannot be estimated. The current study is a first effort to correct this deficit. We have developed an algorithm, based on recurrent correlation, that computes the correlation between measurement domains (e.g., head motion and F0) as a time-varying function. Using this method regions of high or low correlation, or of rapid transition (e.g., from high to low) can be associated with visual and auditory events. The analysis of the time-varying coupling of multi-modal events has implications for speech planning and synchronization between speaker and listener.

1pSC10. Modeling lexical distinctiveness as a function of natural variations in segmental speechreading ability. Edward Auer, Jr (Dept. of Speech-Lang.-Hearing, Univ. Kansas, 1000 Sunnyside Ave., Lawrence, KS 66045)

The ability to perceive spoken language on the basis of optical signals alone (speechread) is known to be widely distributed. Previous computational modeling studies investigating sources of speechreading variability (Auer & Bernstein, 1997) demonstrated that small variations in available segmental distinctiveness were associated with large differences in lexical uniqueness. However, these previous studies modeled the variation in available segmental distinctiveness algorithmically and not on the basis of naturally occurring variation. In the current study, behavioral experiments were performed to measure the available segmental distinctiveness for perceivers differing in speechreading ability. Twelve hearing participants, chosen to systematically vary in sentence-level speechreading ability (6 high and 6 low), identified consonants and vowels presented in consonant-vowel or consonant-vowel-consonant nonsense syllables. The two participant groups differed in vowel identification accuracy, but not in consonant identification accuracy. The groups also differed in the patterning of confusions between segments. Results will be presented from computational experiments performed to assess the effects of variation in the available segmental distinctiveness on the uniqueness of words in the language. The results will be discussed in terms of potential sources of individual variability in speechreading. [Work supported by a NIH/NIDCD R01DC04856.]

1pSC11. The effects of visual-only vowel identification training on speech recognition for adults with hearing loss. Carolyn Richie (Dept. of Commun. Disord., Butler Univ., 4600 Sunset Ave., Indianapolis, IN 46208, crichie@butler.edu)

Speechreading, the use of visual cues to speech, may provide benefit for listeners with hearing loss in everyday communication beyond that achieved with hearing aids alone. Recent work examined the effects of a novel approach to speechreading training, for normal-hearing listeners tested in masking noise designed to simulate a hearing loss [C. Richie and D. Kewley-Port, J. Acoust. Soc. Am. **117**, 2570 (2005)]. That study showed improvements in sentence-level speech recognition for listeners trained to speechread vowels under audiovisual conditions, compared to listeners trained to speechread consonants and untrained listeners. The present study examined the effects of vowel identification speechreading training on sentence-level speech recognition, for adults with hearing loss. Computer-based training and testing was administered visual-only, in order to determine if training in the visual modality alone leads to gains in speech recognition. Results showed that vowel identification training under visual-only conditions led to improved vowel recognition, but that improved vowel identification did not generalize to increased sentence-level speech recognition. These findings suggest that some acoustic information, however minimal, may be beneficial during speechreading training and required for improved sentence-level speech recognition by adult listeners with hearing loss. [Work supported by Butler University HAC Grant No. 027096.]

1pSC12. Auditory-visual speech perception: What isolated articulators contribute. Alexis N. Bosseler, Patricia K. Kuhl (Inst. for Learning and Brain Sci., Box 357988, Univ. of Washington, Seattle, WA 98195), Dominic W. Massaro (Univ. of California, Santa Cruz, Santa Cruz, CA 95064), and Andrew N. Meltzoff (Univ. of Washington, Seattle, WA 98195)

The present study investigated the effects of viewing specific articulators on auditory-visual speech perception. We used a jaw-only, lips-only, jaw+lips, and full face visual stimulus paired with clear versus degraded auditory stimuli in an expanded factorial design. The unimodal and bimodal speech syllables were identified as one of five possible response alternatives (/ba/, /da/, /tha/, /bda/ or /bga/). The results indicate that both visual and auditory sources of information influenced syllable identifica-

tion under the lips alone, lips+jaw alone, and whole face conditions. Performance was better described by the fuzzy logical model of perception than by an auditory dominance model or a weighted-averaging model. Overall, the results are consistent with the idea that although there may be differences in the information provided (which reflect the amount of information each articulator contributes to perception), the way in which the speech information is processed remains invariant. Work funded for the research was provided by a National Institute of Health Training in Speech and Hearing Sciences grant to the University of Washington's Speech and Hearing Sciences Department and by a National Science Foundation Science of Learning Center grant to the University of Washington's LIFE Center.

1pSC13. Dynamic cortical imaging of speech compensation for auditory feedback perturbations. John Houde, Srikantan Nagarajan, and Theda Heinks-Maldonado (Depts. of Otolaryngol. and Radiol., Univ. of California, 513 Parnassus Ave., HSE800, San Francisco, CA 94143, houde@phy.ucsf.edu)

Understanding how auditory feedback is processed and used during speech production is a long-standing issue, one that has classically been investigated by looking at how altering auditory feedback affects speech. Recently, the advent of functional neuroimaging methods has also allowed examination of how producing speech affects the neural processes serving auditory perception. Here, whole-head magnetic source imaging (MSI) was used to monitor cortical activity as speakers responded to brief perturbations of the pitch or amplitude of their speech. Prior studies have shown that such perturbations cause compensatory responses in speech motor output. MSI recordings were acquired at the onset of feedback perturbations as the subject continuously vocalized and also when the subject passively listened to recordings of the perturbed feedback. A response to the perturbations was found around auditory cortex between 100 and 400 ms postperturbation that was enhanced while subjects vocalized, as compared to when they passively listened. The size of this response enhancement was a significant predictor of how much a subject compensated for the feedback perturbation. Time-frequency optimized adaptive spatial filtering was also used to localize cortical activations that correlated significantly with degree of compensation across subjects. [Work supported by NSF grant BCS-0349582 and NIH/NIDCD grant R01-DC006435.]

1pSC14. Does the production of speech necessarily rely on auditory feedback? Stacie K. Byrne, Ingrid S. Johnsrude, and Kevin G. Munhall (Dept. of Psych., Queen's Univ., 62 Arch St., Kingston, ON., K7L 3N6)

Talkers show sensitivity to a range of perturbations of auditory feedback (e.g., manipulation of vocal amplitude, fundamental frequency, formant frequency). When the auditory concomitants of speech (feedback) are perturbed, talkers compensate in order to preserve accurate productions. These effects demonstrate the role of sensory feedback in articulatory control and have been used to study the role of prediction in motor planning. In this study, 15 subjects spoke a monosyllable ('head') and the formants in their utterances were shifted in real time using a custom signal processing system and were fed back over headphones with an imperceptible delay. The first and second formants were altered so that the auditory feedback matched the subjects' productions of 'had'. Despite explicit instruction to ignore the feedback changes and maintain consistent production of 'head', subjects produced a robust compensation similar to the compensatory behavior shown by naive subjects. In general, subjects altered their vowel formant values in a direction opposite to the perturbation, as if to cancel its effects. These results suggest that compensation in the face of formant perturbation is automatic and obligatory. Results will be discussed in terms of the representations that guide speech motor control. Work supported by NIDCD and NSERC.

1pSC15. Activity in regions sensitive to auditory speech is modified during speech production: fMRI evidence for an efference copy. Zhuo Zheng, Ingrid Johnsrude, and Kevin Munhall (Dept. of Psych., Queen's Univ., 62 Arch St., Kingston, Canada, K7L 3N6)

Models of speech production postulate that, in order to facilitate rapid and precise control of articulation, the predicted auditory feedback is sent to the auditory system to be compared with incoming sensory data. If this is so, an "error" signal may be observed when the predicted auditory feedback and the sensory consequences of vocalization do not match. We used event-related fMRI to look for the neural concomitants of such an error signal. In two conditions volunteers whispered "ted." In one of these, voice-gated noise was used to mask the auditory feedback, which should result in an error signal. Two other conditions were yoked to the production conditions (either clearly heard or masked), but were listen-only and therefore no error signal would be expected. We acquired whole-brain EPI data from 21 subjects using a fast-sparse design. Activity in the superior temporal gyrus bilaterally was significantly greater for clear than masked speech during the listen-only trials and significantly higher for masked than for clear speech in the production trials. This crossover interaction indicates that speech production results in corollary discharge in the auditory system and furthermore suggests that this corollary discharge reflects expectations about the sensory concomitants of speech acts.

MONDAY AFTERNOON, 4 JUNE 2007

SALON B, 1:00 TO 3:00 P.M.

Session 1pSPa

Signal Processing in Acoustics: Classification and Interpretation of Acoustic Signals

George E. Ioup, Chair
Univ. of New Orleans, Dept. of Physics, New Orleans, LA 70148-0001

Contributed Papers

1:00

1pSPa1. Coarse classification of acoustic signals using temporal and spectral characteristics. Daniel J. Mennitt, Marty Johnson, and James Carneal (Vib. and Acoust. Lab., Virginia Tech., 143 Durham Hall, Blacksburg, VA 24061-0238, dmennitt@vt.edu)

Sound can be classified in many ways on many levels of detail. Coarse classification is especially desirable in situations where the expected type of acoustic sources is unknown. A reliable method is proposed to classify acoustic signals as impulsive, tonal, or broadband (e.g., gunshot, engine, and vehicle tire noise) based on statistics of the signal in the time and frequency domains. Algorithms are presented that efficiently calculate an impulsive and tonal metric, which is used to differentiate multiple acoustic sources in a noisy environment. This is applicable to multitarget tracking where sequential measurements of a target's location must be associated to propagate the track; the acoustic attribute discriminates against targets with similar kinematic quantities. Elementary signals are subjected to the proposed classification algorithm to determine thresholds of the categories. Finally, the algorithms are shown to identify various sources from measured field data.

1:15

1pSPa2. Robust ground vehicle classification using the modulation spectrogram. Geok Lian Oh (DSO Natl. Labs., 20 Science Park Dr., Singapore 118230)

The performance of a vehicle classification system using acoustic emissions is affected by acoustic interferences such as wind noise. In this paper, we propose using the modulation spectrogram as features for classification. This is done by extraction of the temporal structure in the vehicular sounds that appear as amplitude modulations in the frequency band corresponding to the harmonic structure present in the vehicular sounds. The amplitude modulations are processed in subband channels following critical band frequency analysis. Our results of the visual displays of vehicle sounds produced with the modulation spectrogram are relatively stable in the presence of wind noise. This suggests that the use of the modulation representation is another method for improving the robustness of vehicle classification system. In the paper, we apply the suggested approach to classify between four classes of vehicles and obtain reasonably good classification rate even for data contaminated with interferences.

1:30

1pSPa3. Towards improving the target recognition using spectral and cepstral features. M.H. Supriya, and P.R. Saseendran Pillai, (Dept. of Electron., Cochin Univ. of Sci. and Technol., Cochin 682 022, India)

The identification of various noise sources in the ocean, which are generating noises that are diverse in characteristics, carried out using power spectral estimation has only limited success rates. It has been found that the errors associated with the identification process can be reduced, to some extent, if the information gathered from the power spectral estimation is supplemented with those gathered from cepstral analysis. A prototype system for identifying the targets from the target signatures generated from the noise waveforms emanating from the noise sources in the ocean has been implemented by judiciously combining the features extracted from the spectral estimation and cepstral analysis. It has been observed that the MFCC for various records vary over a range of values and in order to improve the performance of the identifier, the optimum set of values for the cepstral coefficients are synthesized by a technique quite frequently referred to as vector quantization. In this paper, the LBG (Linde, Buzo, Gray) design algorithm for vector quantization by trimming the cepstral coefficients to the nearest value has been adopted for finding out the optimum match.

1:45

1pSPa4. Detecting and identifying non-stationary signals. Duncan P. Williams and Aled T. Catherall (Dstl Physical Sci., Dorchester, DT2 8WX, UK)

Detecting and identifying non-stationary signals is well known to be important in many sonar applications. While great strides have been made in the last 60 years of underwater sonar, every effort is still needed to detect and then identify whatever glimpses in the ocean can be found. The traditional method for analyzing such signals, which contain time-varying frequencies, is the short-time Fourier transform (STFT) or lofargram. However, neither the STFT nor other methods, such as wavelet transforms and Wigner-Ville distributions, that have found favor in speech processing, are always well-suited to non-stationary signals. The story remains the same out of the water too. The fractional Fourier transform (FrFT) has emerged in recent years as a powerful tool for analyzing complicated non-stationary signals. As an example, we first apply the FrFT, alongside other time-frequency methods, to examine real bat chirps, marine mammal

soundings, and speech signals. A new short-time FrFT, which accounts for signals with multiple non-stationary components, is then demonstrated, allowing us the flexibility to track several different signals at the same time. The remainder of the paper outlines some of the classification techniques, which complement the output from the short-time FrFT, trying to identify non-stationary signals of every kind.

2:00

1pSPa5. A comparison study of various techniques for blind estimation of reverberation time. Joshua Fialkoff and Ning Xiang (Graduate Prog. in Architectural Acoustics, Rensselaer Polytech. Inst., Troy, NY 12180)

Reverberation in a room often adds a desirable coloration to a perceived sound, but research shows that one's ability to correctly interpret speech degrades with increased reverberation. Recently, various techniques for blind estimation of reverberation time have been suggested. Techniques such as these could prove to be indispensable tools not only for architectural acousticians, but communication engineers and audiologists as well. In this study, various techniques for blind estimation of reverberation time are implemented in the MATLAB environment. Each technique is tested using various real and simulated signals to determine its respective efficacy.

2:15

1pSPa6. Decomposition of the time-reversal operator in the Rayleigh limit. Jean-Gabriel Minonzio, Claire Prada, and Mathias Fink (Laboratoire Ondes et Acoustique, Université Denis Diderot Paris 7, UMRS CNRS 7587, ESPCI, 10 rue Vauquelin 75231 Paris cedex 05, France)

Acoustic scattering analysis is an important tool in acoustic imaging and characterization with applications in non-destructive testing, medical imaging or underwater acoustics. The DORT method consists of the study of the time-reversal invariants which correspond, for a given transducers array, to the eigenvectors of the time-reversal operator or to the singular vectors of the array response matrix **K**. The decomposition of the scattered pressure in normal modes of vibrations allows the determination of the theoretical time-reversal invariants for different two-dimensional elastic objects such as cylinders or tubes. The N dimension problem, where N is the number of transducers, is reduced to a problem which dimension is about $2ka+1$, where a is the scatterer radius and k the wave number in the surrounding fluid. This approach provides analytical expressions of the two first singular values and vectors in the Rayleigh limit, i.e. $2a$ smaller than the resolution cell. These results are verified experimentally and in

good agreement with the original point of view: for a small scatterer, there is one dominant singular value associated with the singular vector which focuses isotropicaly on the scatterer. Some of these results can be generalized to 3-D objects such as spheres.

2:30

1pSPa7. Role of the Richardson extrapolation in the evaluation of acoustic data: Theory. Allan J. Zuckerwar (Analytical Services and Mater., Inc., 107 Res. Dr., Hampton, VA 23666, a.j.zuckerwar@larc.nasa.gov)

In many acoustical applications the method of least squares is a popular method for fitting experimental data to known physical laws or empirical functions. There are occasions when the method falls short of attaining the goals of the experimenter; for example, when estimating the value of a dependent variable beyond the range of the data (sound speed at zero frequency), fitting to a physical law described by an infinite series (acoustical virial coefficients), or fitting to a semi-empirical relationship when the governing physical law is unknown (acoustic ground impedance). Following a description of the Richardson extrapolation and the methodology for its implementation, it will be shown to resolve many of the shortcomings of a least squares analysis in the evaluation of acoustic data.

2:45

1pSPa8. Role of the Richardson extrapolation in the evaluation of acoustic data: Applications. Allan J. Zuckerwar (Analytical Services and Mater., 107 Res. Dr., Hampton, VA 23666, a.j.zuckerwar@larc.nasa.gov)

Three acoustical applications of the Richardson extrapolation procedure are described, where in each case it is not possible to make a measurement at the zero value of the independent variable: (1) the vibrational relaxation time of nitrogen in the limit to zero humidity; (2) the speed of sound of methane in the limit to zero pressure (and, subsequently, its second and third acoustical virial coefficients); (3) the acoustic impedance of a grass covered field in the limit to zero frequency. In example (1), a linear regression extrapolated to zero humidity yields a physically untenable value of the reciprocal relaxation time. In example (2), in which the number of terms in the governing physical law is infinite, the results of a least-squares best-fit depends on the selected number of terms in the representative polynomial (truncation error). In example (3) the governing physical law is unknown, but semi-empirical models have a low-frequency limit that precludes the prediction of the impedance of the ground in the limit to zero frequency. The Richardson extrapolation procedure is shown to resolve these difficulties.

Session 1pSPb

Signal Processing in Acoustics: Signal and Array Processing of Acoustic Data

Paul Hursky, Chair
Heat Light and Sound Research, 12730 High Bluff Dr., San Diego, CA 92130

Contributed Papers

3:15

1pSPb1. Passive reverberation nulling for target enhancement.
Heechun Song, William S. Hodgkiss, William A. Kuperman, Karim Sabra, Tuncay Akal (MPL/SIO, 9500 Gilman Dr., La Jolla, CA 92093-0238), and Mark Stevenson (NURC, La Spezia, Italy)

Echo-to-reverberation enhancement has been previously demonstrated using time reversal focusing when knowledge of the channel response between a target and the source array element is available. In the absence of this knowledge, direct focusing is not possible. However, active reverberation nulling still is feasible given observations of reverberation from conventional source array transmissions. For a given range of interest, the response between the source array elements and the dominant sources of boundary reverberation is provided by the corresponding reverberation from this range. Thus, an active transmission can be projected from the source array which minimizes the energy interacting with the boundaries at a given range while still ensonifying the waveguide between the boundaries. As an alternative, here we describe a passive reverberation nulling concept. In a similar fashion, the observed reverberation defines the response between the source array elements and the dominant sources of boundary reverberation at each range and this is used to drive a range-dependent sequence of projection operators. When these projection operators subsequently are applied to the received time series, reverberation can be suppressed. The improvement in target detectability is demonstrated using experimental data with an echo repeater simulating the presence of a target in shallow water.

3:30

1pSPb2. Single and multiple channel estimation using the time-updated recursive least squares algorithm. Weichang Li and James Preisig (Appl. Ocean Phys. and Eng. Dept., Woods Hole Oceanogr. Inst., M.S. 9, Woods Hole, MA 02543)

A time-updated recursive least squares algorithm (TURLS) is presented, analyzed, and applied to the problem of estimating the time-varying impulse response in both single and multiple channel cases. The TURLS algorithm may be viewed as a least squares extension of the EM channel estimation algorithm [W. Li and J. C. Preisig, ICASSP, IV 313IV, 316 (2005)], in the sense that it jointly estimates the channel impulse response and its dynamic parameters by minimizing a least squares error instead of maximizing the likelihood function. Based on the intuition that the sequence of channel estimates contains information regarding the true channel dynamics, the algorithm estimates the parameters from the sequence of channel estimates. Analysis of the algorithm based on a first-order AR channel model shows that the resulting channel estimate is a second-order approximation whose poles converge as the parameter estimate improves. In addition, it is shown that with exponential weighting, the resulting parameter estimate may be biased at low SNR level. The algorithm is applied to experimental data and results with both single channel and multiple channels are demonstrated. [Work supported by ONR Ocean Acoustics.]

3:45

1pSPb3. Fluctuation-based processing meets Hilbert transform.
Ronald A. Wagstaff (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, rwagstaf@olemiss.edu)

Fluctuation-based processing (FBP) exploits fluctuations in amplitude and phase to achieve signal processing gains. Three consecutive fast Fourier transforms (FFTs) are required to obtain an exploitable FBP phase acceleration parameter. In addition, at least three values of this parameter are required to identify signal presence and/or to achieve the full potential of FBP. An example of interest is an acoustic probe on an unmanned air vehicle (UAV) flying among the atmosphere boundary layer, where the turbulence noise persists a significant fraction of time. In such a case, the FFT time periods have a high level of confidence of being contaminated by turbulent flow noise. Furthermore, the average of several FFTs will have a much higher level of confidence of being contaminated. Contaminated results generally have no value for identifying signal presence. One solution for this problem is a Hilbert transform, which provides temporal resolution in phase and amplitude at the sample rate, e.g., 0.001s. With such high temporal resolution, small time segments of relatively uncontaminated phase and amplitude can be automatically identified by the FBP algorithms and successfully exploited for gain. This will be demonstrated with measured air-borne acoustic sensor data. [Work supported by AR-DEC.]

4:00

1pSPb4. Eye array dereverberation by corner placement. Hedayat Alghassi, Shahram Tafazoli, and Peter Lawrence (Dept. of Elec. and Computer Eng., Univ. of British Columbia, Vancouver, BC, Canada, hedayata@ece.ubc.ca)

A new signal processing algorithm accompanied with a novel array structure for sound source localization (SSL) in three-dimensional spaces was presented [Hedayat Alghassi *et al.*, J. Acoust. Soc. Am. **120**(5) (2006)]. This methodology has some analogy to the eye. The microphone array's proper placement is particularly important in enclosed areas, where reverberation is the dominant problem. It will be demonstrated that by placing the eye array's symmetry axis along the boresight axis of an upper orthogonal trihedral corner in an enclosed area, the adverse effect of early reverberation can be substantially reduced. This enhancement is the consequence of both the retroreflection property of the orthogonal trihedral corners plus the eye array's insensitivity to back reflections. There is a pseudo-coincidence between the estimation cells and the source direction in the eye array SSL method. Therefore, the majority of the early reflections from the trihedral corner strike the estimation cells in the direction opposite to the sound source, where the sensitivity is lowest. This approach not only reduces the adverse effect of reverberation on the SSL accuracy, but also creates an approximately flat error for all source directions. Experiments confirmed that this method achieves a reasonable accuracy improvement.

4:15

1pSPb5. Impact of underwater noise environment on coherent array processing of global positioning system (GPS) sonobuoy fields. Abdalla Osman, Aboelamgd Nourledin, Jim Theriault, and Scott Campbell (Dept. of Geomatics Eng., Univ. of Calgary, 2500 University Dr., N.W. Calgary, AB, Canada T2N 1N4)

Global positioning system (GPS) sonobuoys are widely used in underwater target localization. Most of the present underwater target tracking utilizing GPS sonobuoys rely on individual processing of the data provided by each sonobuoy. Coherent array processing of GPS sonobuoy fields was recently suggested in order to improve the overall system performance by proper combination of sonobuoys received signal. Therefore, it is essential to study the impact of underwater noise environment on target localization when an array of GPS sonobuoys is utilized. This paper introduces a complete analysis of the effect of different types of noisy environments. The overall system performance is evaluated for each environment for both individual sonobuoy processing and coherent array processing. Simulation was used to examine the effect of both white Gaussian noise and correlated noise on the overall system performance in the presence of two sources of interference at variable signal to noise ratios (SNR). Each sonobuoy signal is de-noised by wavelet analysis prior to processing multiple sonobuoy signals by beamforming to provide accurate bearing information for the target. The results discuss the impact of wavelet de-noising and the array processing in improving the overall system accuracy for different noisy environments at different SNRs.

4:30

1pSPb6. An investigation of the use of chaotic signals in active sonar. Alan Fenwick (Air Vehicles, Concepts & Structures, QinetiQ, Farnborough, GU14 0LX, UK)

Many natural effects are now known to be chaotic rather than random, i.e., are generated by simple non-linear systems. The paper reports the outcome of a theoretical and experimental study to assess whether signals based on chaotic time series offer advantages in active sonar. It is noted that for conventional detection, the detailed structure of the signal is unimportant. Any advantage will come from novel effects due to interactions with the target or the environment. An overview is given of apparently novel behavior in echoes from a resonant target which were first noted in simulations and then observed in laboratory ultrasonic measurements. An explanation in terms of linear systems theory is outlined. The limitations which inhibit the use of advanced non-linear techniques for detection will be discussed and illustrated with reference to the results of analyzing data measured in an outdoor acoustic test facility.

4:45

1pSPb7. Estimating the state of a one-dimensional waveguide. Edgar Berdahl and Julius Smith III (Ctr. for Comput. Res. in Music and Acoust. (CCRMA), The Knoll, 660 Lomita, Stanford, CA 94305, eberdahl@ccrma.stanford.edu)

Methods for estimating the state of a one-dimensional waveguide are analyzed. The state is parameterized in terms of right-going and left-going traveling waves for applications in reflectometry and control. Traditional methods estimate the localized state at a single position in the waveguide. In many cases, the history of a simple estimator may serve as time-delayed or time-advanced estimates for the remainder of the waveguides state. The decomposition method estimates the traveling wave components at a point using integration and a spatial gradient approximation. The delay method, which involves implementing a single delay accurately, is shown to be equivalent to the causal Wiener filter least-squares optimal estimator. If sensors are spaced irregularly, then an optimization problem typical of beam forming applications may be solved to find an FIR filter-based localized estimator. On the other hand, when the estimator needs to estimate the entire waveguide state quickly, then either additional sensors or computation cycles are required. The Kalman filter requires much more computation in general, but with the help of a digital waveguide model, it can provide complete state estimates with less than one sample of delay using as few as one sensor. [Work supported by the Wallenberg Global Learning Network.]

5:00

1pSPb8. Distributed multi-microphone (dmm) classification. Marek B. Trawicki, Michael T. Johnson (College of Eng., Marquette Univ., P.O. Box 1881, Milwaukee, WI 53201-1881, marek.trawicki@marquette.edu), and Tomasz S. Osiejuk (Adam Mickiewicz Univ., 61-614, Poznan, Poland)

Over the past several decades, research in signal enhancement and speech recognition has concentrated on single channels and microphone arrays. Whereas single channels require subjects who are relatively close to the microphone, microphone arrays require close spacing and *a priori* knowledge of the geometry. In contrast to those stringent conditions, distributed multi-microphones (DMMs) can be utilized for situations that require the microphones that are positioned far away from the subjects with possibly unknown wide-spacing and configurations such as in meeting rooms or the wild. As opposed to performing recognition through microphone selection, feature integration, or likelihood combination, the proposed work focuses on processing the DMM signals to diminish the effects of ambient noise and form one optimal signal before passing it into the recognizer through two methods: weighted sum of distances and weighted sum of signal powers. Song-type classification experiments are presented on eight-channel Norwegian Ortolan Bunting (Emberiza Hortulana) vocalizations over a microphone range of 1 to 206 m from the reference on 1620 recordings using cepstral coefficients, energy, and time derivatives features. Based on the results, the two methods achieve accuracies of 91.2% (signal powers) and 94.4% (distances) against 90.7% (closest microphone) on 804 test exemplars divided across four song types.

Session 1eID

Interdisciplinary: Tutorial Lecture: Musical Acoustics: Science and Performance

James W. Beauchamp, Chair

Univ. of Illinois Urbana-Champaign, School of Music, Dept. Electrical and Computer Engineering, 114 W. Nevada, Urbana, IL 61801

Chair's Introduction—7:00

7:05

1eID1. Musical acoustics: Science and performance. Uwe J. Hansen (Indiana State Univ., Terre Haute, IN 47809-9989) and Jerry Floor (Salt Lake City Jazz Orchestra, Salt Lake City, UT)

Musicians of the Salt Lake City Jazz Orchestra and their director, Jerry Floor, are staples of the Jazz Scene in the Intermountain West. This evening, rather than limiting your interest to the scientific elements of musical acoustics, will lift your spirits with the rich, exciting sounds of this wonderful ensemble. Before a particular instrument family is featured in a performance, Uwe will simply help you understand why, for example, a clarinet sounds different than a trumpet. The basic science of each instrument group will be introduced before you have an opportunity to fill your ears with the sound of that family. You will experience each family, Strings, Brasses, Woodwinds, Piano, and Percussion through the lush sounds of the Salt Lake City Jazz Orchestra. At the same time you will gain some insight into the nature of tone production and tone quality for each instrument, you will actually see some wave-shapes and some spectra. The evening will conclude with a number of full ensemble performances of some of your favorite music.

Session 2aAA

Architectural Acoustics: An Acoustical Tour of Three Salt Lake City Concert Halls

Timothy W. Leishman, Chair
Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602

Chair's Introduction and Tour Orientation

This session has been organized as a walking and traveling tour of three concert halls in the Salt Lake Valley. The halls are of a variety of sizes, with both fixed and variable acoustics. A 45-minute tour and technical discussion will be led at each venue by the acousticians and architects who contributed to its design. Pianist Larry Jackstien and other musicians will accompany the tour group and perform briefly in each hall. The halls and speakers will include the following:

Abravanel Hall (2,768 seats). Acoustician: Cyril M. Harris. Architect: Franklin T. (Frank) Ferguson of FFKR Architects.

Libby Gardner Concert Hall (680 seats). Acoustician: L. Gerald (Jerry) Marshall of Marshall/KMK Acoustics [represented by Richard K. ("Jim") Fullmer of Spectrum Engineers]. Architect: Franklin T. (Frank) Ferguson of FFKR Architects.

Waterford School Concert Hall (1200 seats). Acoustician: Richard K. ("Jim") Fullmer of Spectrum Engineers. Architects: Joshua R. Vel and Kyle S. Taft of MHTN Architects.

Following the chair's introduction, there will be a short 10-minute walk (two city blocks) from the meeting hotel to Abravanel Hall. After the presentation at Abravanel Hall, chartered bus transportation will be provided to Libby Gardner Concert Hall and Waterford School Concert Hall. The bus should return to the meeting hotel before 12:30 p.m. Box lunches will be provided for a minimal charge to those who wish to eat en route to the hotel. Because space is limited, advanced registration is required. Register by sending your name and contact information to Tim Leishman [e:mail: tim_leishman@byu.edu or fax: (801) 422-0553]. You will be contacted regarding your lunch preferences.

Session 2aAB

Animal Bioacoustics: Paleohearing

Richard R. Fay, Chair
Loyola Univ., Parmly Hearing Inst., 6430 N. Kenmore, Chicago, IL 60626

Chair's Introduction—8:30

Invited Papers

8:35

2aAB1. Evolution of hearing specializations in teleost fishes. Christopher B. Braun (Dept. of Psych., Hunter College of the City Univ. of New York, 695 Park Ave., New York, NY 10021, cbraun@hunter.cuny.edu) and Terry Brande (Loyola Univ. of Chicago, Chicago, IL 60626)

Among teleost fishes, the spectacular diversity of body forms, ecotypes, and behavioral repertoires is matched by a tremendous diversity of hearing mechanisms and, presumably, auditory function. Morphological specializations with the potential to enhance hearing have arisen as many as 20 times within teleosts. These specializations range from modest modifications of the buoyancy-regulating gas bladder to wholesale reorganization of the rostral vertebral column and posterior neurocranium. Since many of these specializations involve vertebral elements and/or cranial modifications, it is possible to recognize enhanced hearing in fossil taxa. Other specializations, like the rostral swimbladder extensions of some holocentrids, appear to provide similar, but more modest, enhancements. In other cases, however, rostral swimbladder extensions may only increase the upper frequency limit (as in some scaeinids) or somehow underlie improved spectral resolution (in notopterids). The functions of these and many other specializations await more complete functional characterization. For instance, the number, linkages, and robustness of Weberian ossicles vary considerable within Otophysi, and it is clear that hearing abilities differ widely, too. How and why individual Weberian apparatus morphologies confer differences in hearing abilities remains a critical issue in studies of fish hearing.

2aAB2. Audiogram, body mass, and basilar papilla length: Correlations in birds and predictions for extinct archosaurs.
Robert Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, dooling@psyc.umd.edu), Otto Gleich (Univ. of Regensburg, Regensburg, Germany), and Geoffrey Manley (Lehrstuhl für Zoologie, Technische Universität München, Garching, Germany)

The inner ear in the group of archosaurs (birds, crocodilians, and extinct dinosaurs) shows a high degree of structural similarity, enabling predictions of their function in extinct species based on relationships among similar variables in living birds. Behavioral audiograms and morphological data on the length of the auditory sensory epithelium (the basilar papilla) are available for many avian species. By bringing different data sets together, we show that body mass and the size of the basilar papilla are significantly correlated, and the most sensitive frequency in a given species is inversely related to the body mass and the length of the basilar papilla. We also demonstrate that the frequency of best hearing is correlated with the high-frequency limit of hearing. Small species with a short basilar papilla hear higher frequencies compared with larger species with a longer basilar papilla. Based on the regression analysis of two significant correlations in living archosaurs (best audiogram frequency versus body mass, and best audiogram frequency versus papillar length), we suggest that hearing in large dinosaurs was restricted to low frequencies with a high-frequency limit below 3 kHz.

2aAB3. The evolution of the mammalian middle ear.
John J. Rosowski (Eaton-Peabody Lab., Mass. Eye & Ear Infirmary, 243 Charles St., Boston, MA 02114)

The transition from water-to-land in vertebrate evolution precipitated physical differences between sound transmission in the new surround (air) and the water-filled inner ear. These physical differences led to the development of the vertebrate middle ear. The old idea that the middle ear first developed in amphibians, was refined in reptiles and birds, and was further refined in mammals has been contradicted by the available paleontologic evidence which clearly points to the independent development of the tympanic-ossicular middle ear in multiple vertebrate lines, including amphibians, reptiles, and mammals. Indeed, mammalian and reptilian evolutionary trees diverged over 300 million years ago at a time well before the development of either the reptilian or mammalian middle ear. The evidence on the evolutionary development of the mammalian middle ear will be reviewed and put in context of the development of the reptilian and bird middle ears. Attempts to describe the function of the early mammalian ear will also be made through comparisons with the function of ears of similar structure in living animals. The significant influence of the evolving inner ear will be a major point of discussion. [Work supported by the NIDCD.]

2aAB4. Searching among felids for evolution's influences on middle ears.
William T. Peake, John J. Rosowski, and Helen C. Peake (Eaton-Peabody Lab., Massachusetts Eye & Ear Infirmary, 243 Charles St., Boston, MA 02114)

The 36 cat species display moderate structural variations that may be adaptive, involving species ethology, size, and lineage. Measurements of skull dimensions, acoustic responses in live and cadaver ears, and auditory brain-stem responses in anesthetized specimens have been collected. Of special interest are interspecies variations in the acoustic properties of the felid middle-ear air space, which has two coupled cavities separated by a bony septum. The acoustic effect of the cavities is to decrease middle-ear response uniformly for low frequencies, and to produce a narrow-band reduction at a mid frequency determined by the cavities configuration. We address two issues: (1) The volume of the cavities increases with species size. A physics-based model shows that the cavity-related loss in low-frequency middle-ear transmission decreases from 10 dB for the smallest species to 1 dB in lions. Is this size-dependent variation adaptive? (2) The configuration of the cavities-dividing septum is unusual in eight species, all of which have relatively wide facial skulls and live in open habitats; as these species represent six of nine lineage groups, this constellation of habitat, skull, and ear features apparently evolved several times. Can we hypothesize a testable, adaptive mechanism? [Work supported by NIH and NSF.]

2aAB5. Evolution of mammalian hearing.
Henry E. Heffner and Rickye S. Heffner (Dept. of Psych. MS 948, Univ. of Toledo, 2801 W. Bancroft St., Toledo, OH 43606, Henry.Heffner@utoledo.edu)

The comparative study of mammals reveals systematic differences in their hearing abilities. First, the ability to hear above 10 kHz (which apparently evolved for using pinna cues to localize sound) varies inversely with the size of an animal's head such that smaller animals usually have better high-frequency hearing than larger ones. Second, low-frequency hearing ability is bimodally distributed: Although most mammals can hear below 125 Hz, about one third do not hear much below 500 Hz and probably do not use temporal coding for pitch perception. Third, sound localization acuity varies directly with the width of the field of best vision, indicating that a major source of evolutionary pressure on hearing is to direct the eyes to the source of a sound. Finally, while some mammals do not use the binaural time-difference locus cue and a few others do not use the binaural intensity difference cue, it is likely that all mammals that possess pinnae use pinna locus cues to localize sound. Taken together, these findings suggest that some of the hearing abilities of extinct mammals may be inferred from fossil evidence bearing on their head size, eyes, and the presence of pinnae.

10:30

2aAB6. The molecular basis for auditory system evolution. Bernd Fritzsch, Feng Feng, and David Nichols (Biomed. Sci., Creighton Univ., Omaha, NE 68178, Fritzsch@Creighton.edu)

Evolution of the auditory system requires coordinated changes in three interrelated systems: the ear, sensory neurons, and brainstem. The auditory periphery evolved by building on the capacity of vestibular hair cells to respond to mechanical stimulation, then segregating a distinct sensory patch dedicated to processing auditory information and evolving accessory structures. Auditory ganglion neurons share much of the molecular basis of their development with vestibular neurons and are, therefore, likely derived from vestibular neurons, possibly through the expression of specific genes such as the zinc finger protein Gata3. In mammals, Gata3 is selectively associated with spiral sensory neurons. Hair cells, sensory neurons, and central auditory nuclei development depends on only two bHLH genes, Atoh1 and Neurog1. Atoh1 null mutant mice never differentiate hair cells and show an almost complete loss of cochlear nuclei. In contrast, Neurog1 null mice form no sensory neurons. Overall, the ear hair cells, sensory neurons, and the brainstem auditory nuclei show molecular conservation of critical regulatory genes, embedded in an organ-specific molecular context which results in the modification of the developmental pathways governed by these conserved molecules to serve either hair cell or cochlear nucleus neuron differentiation in the case of Atoh1.

Contributed Papers

2aAB7. Representation of particle motion in the developing midbrain. Andrea Simmons, Seth Horowitz, and Leslie Tanyu (Dept. of Psych. and Neurosci., Brown Univ., Providence, RI 02912)

In underwater organisms such as tadpoles, both particle motion and pressure components provide important information about ambient sound sources. Using both anatomical and physiological techniques, we examined the representation of these components in the developing midbrain. Afferents from the lateral line and from the saccule both project to the medial vestibular nucleus in the medulla which, in turn, sends projections to the laminar nucleus of the torus semicircularis in the midbrain. Extracellular recordings from the laminar nucleus of tadpoles from early larval through metamorphic climax stages show robust responses to particle motion, but across a more constrained frequency range than observed in the medulla. Combined with earlier data showing good responses to the pressure component of sound sources, these data suggest that the laminar nucleus may be a site of integration of particle motion and pressure sensitivity.

2aAB8. Binaural interaction in the medulla and midbrain of toadfish (*Opsanus tau*). Richard Fay and Peggy Walton (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

Responses of cells in the toadfish medulla (descending octaval nucleus: DON) and midbrain (torus semicircularis: TS) were studied to investigate binaural interaction and processing. Normally, the two ears of fish cannot be stimulated independently. A method was developed to temporarily inactivate one ear by slightly displacing the saccular otolith on one side (tipping) to change its orientation in space and, therefore, alter the responsiveness of the hair cells. Brain cells were evaluated for directional characteristics and frequency response (1) before otolith tipping, (2) with the otolith tipped, and (3) post-tipping. For DON cells ($n=14$), contralateral saccular otolith tipping most often resulted in subtle effects consisting of an overall change in responsiveness ($\pm$ spikes/sec); significant changes in the preferred direction were rare. In the TS ($n=20$), most cells exhibited changes in responsiveness and in directionality. These experiments demonstrate the existence of excitatory and inhibitory binaural interactions in the medulla and midbrain and show rather complex and unexpected binaural effects on the responsiveness and directional properties of auditory brain cells. [Work supported by the NIH/NIDCD.]

TUESDAY MORNING, 5 JUNE 2007　　　　　　　　　　CANYON ROOM B, 8:00 TO 9:30 A.M.

Session 2aAOa

Acoustical Oceanography: Oceanographic Variability

Kathleen E. Wage, Chair
George Mason Univ., Dept. of ECE, 4400 University Dr., Fairfax, VA 22030

Contributed Papers

2aAOa1. Signal processing techniques for low-order acoustic modes. Tarun K. Chandrayadula, Kathleen E. Wage (George Mason Univ., 4400 University Dr., MSN 1G5, Fairfax, VA 22030), James A. Mercer, Bruce M. Howe (Univ. of Washington, Seattle, WA 98195), Peter F. Worcester, and Matthew A. Dzieciuch (UCSD, La Jolla, CA 92093)

Internal waves introduce substantial fluctuations in the lowest mode signals recorded during long-range tomography experiments. Lacking a complete random model for the mode signals, tomographers typically resort to averaging the mode signals across receptions to mitigate internal

wave effects. Chandrayadula *et al.* used the 2004 Long Range Ocean Acoustic Propagation EXperiment (LOAPEX) signals and parabolic equation (PE) simulations modeling the LOAPEX environment to derive range-dependent statistics such as temporal mean, temporal covariance, and intermodal correlation of the lower mode signals [Chandrayadula *et al.*, J. Acoust. Soc. Am. **120**, 3062 (2006)]. This talk proposes several statistical signal processing techniques such as likelihood ratio detectors, minimum entropy deconvolution methods, and empirical orthogonal function detectors that are based on the derived statistics, to mitigate internal wave effects. The proposed methods are tested on both the LOAPEX signals and PE simulated mode signals and compared against simple averaging techniques. [Work supported by ONR.]

8:15

2aAOa2. A decade of acoustic thermometry in the North Pacific Ocean: Using long-range acoustic travel times to test gyre-scale temperature variability derived from other observations and ocean models. Brian Dushaw, R. Andrew, B. Howe, J. Mercer, R. Spindel (Appl. Phys. Lab., Univ. of Washington, 1013 N.E. 40th St., Seattle, WA 98105, dushaw@apl.washington.edu), P. Worcester, B. Cornuelle, M. Dzieciuch (Scripps Inst. of Oceanogr., La Jolla, CA), T. Birdsall, K. Metzger (Univ. of Michigan, Ann Arbor, MI), and D. Menemenlis (J.P.L., Calif. Inst. Tech., Pasadena, CA)

Large-scale temperatures in the North Pacific were measured by long-range acoustic transmissions from 1996–2006. Acoustic sources off California and Kauai transmitted to receivers distributed throughout the North Pacific from 1996–1999. Kauai transmissions continued from 2002–2006. Acoustic travel-time data are inherently integrating. This averaging suppresses mesoscale variability and provides an accurate measure of large-scale temperature, subject to the limitations of the ray path sampling. At basin scales, the ocean is highly variable, with significant changes occurring at time scales from weeks to years. The interannual variability is large compared to trends in the data. Willis, *et al.* used objective mapping techniques applied to satellite altimetry and hydrography to derive 0–750 m temperature fields for the global ocean. Travel times equivalent to the measured travel times can be calculated using these fields. The measured and calculated travel times are similar, but also show significant differences. Similar comparisions using travel times derived from the ''Estimating the Circulation and Climate of the Ocean'' (ECCO) model and a high-resolution Parallel Ocean Program (POP) model also show similarities and differences. The ECCO model was constrained by altimetric and profile data by data assimilation, suggesting that the acoustic travel times provide meaningful additional constraints on model behavior.

8:30

2aAOa3. Haro Strait topography (water depths). A. Tolstoy (ATolstoy Science, 1538 Hampton Hill Circle, McLean, VA 22101)

Results have previously shown (most recently at the Honolulu meeting of the ASA, 12/06) that determining important geometric parameters for a single SR path of the Haro Strait test is problematic. In particular, many combinations of parameters will reproduce (via the Collins RAMPE) the observed time-domain arrivals seen in the data. These geometric parameters include water depths at the source and array phones. More recently multiple sources to a single array have still not produced a unique ''solution'' per SR path. This paper will select depths from the reduced yet allowable set of values and attempt to estimate a more global topography for the region via a tomographic method developed earlier.

8:45

2aAOa4. Fluctuations of modal amplitudes due to modes coupling in shallow water. Boris Katsnelson, Valery Grigorev (Voronezh Univ., 1 Universitetskaya sq., Voronezh 394006, Russia), Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), and James Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

In this work, frequency dependence of temporal fluctuations of modal amplitudes is considered. These fluctuations are provided by modes coupling due to perturbation of thermocline layer, for example, as a result of moving soliton (solitons train). In this case, due to the narrowness of the thermocline layer, we have predominating coupling of adjacent modes and

the most significant contribution to intensity fluctuations should be given by a pair of modes. The corresponding numbers of mentioned modes depend on frequency; with increasing frequency numbers of interacting modes increase as well. Both theoretical analysis and experimental results, taken from SWARM95 and SW06, are considered. Reasonable agreement between theoretical calculations and experimental results is demonstrated. [Work was supported by RFBR and ONR.]

9:00

2aAOa5. Spatial and temporal fluctuations of the sound field during shallow water 2006 (SW06) experiment. Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), James Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Boris Katsnelson, and Alexander Tshoidze (Voronezh Univ., Voronezh, 394006, Russia)

During a multi-institutional shallow water experiment in the summer of 2006 (SW06) a series of source-receiver positions were designated to examine the effects of 3D environmental fluctuations on the propagation of low to mid frequency (0.05–3 kHz) broadband acoustic signals. During each scheduled transmission a ship equipped with an acoustic source and a vertical receiver array was positioned at a fixed coordinate determined by the real-time measured position of an internal wave (IW) regime. The passing IW packet in each one of these events was measured by additional research vessels to assess the oceanographic conditions. At the source-receiver ship, the passing IW was observed and its speed and wavelength was measured from surface features using ship's radar within 3-NM radius from the leading edge of the IW on approach. Each event was photographed at 30-s intervals. In addition, thermistor array data were collected simultaneously. A total of 57 passing IW events were recorded during 21 days through which broadband acoustic signals were transmitted at different source-receiver angles with respect to the IW front. This paper presents preliminary data processing of the observed data. [Work supported by ONR-321OA and RFBR.]

9:15

2aAOa6. Observations of phase and travel time variations of normal modes during tropical storms Ernesto and Florence. Gregor Langer, James H. Miller, Gopu R. Potty (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882, g.langer@tu-bs.de), and James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Acoustic normal modes in shallow underwater environments are stretched and compressed due to surface gravity waves and, therefore, experience phase and travel time variations at the receiver. Using the adiabatic approximation, the arrival times of repeated transmissions will wander as a function of time. In addition, the arrivals of fully coupled modes will spread due to the presence of the surface waves [Lynch, Miller, and Chiu, J. Acoust. Soc. Am. **85**, 83–89, (1989)]. Data collected during the Shallow Water 06 experiment on the continental shelf off New Jersey by five single hydrophone receiving units (SHRU) and a vertical line array/horizontal line array were analyzed. Phase-encoded signals from acoustic tomography sources with center frequencies of 224 and 400 Hz were recorded on the receiving systems. The bandpass impulse response of the waveguide was measured repeatedly over a month and the phase and travel time variations of normal modes were determined. The passing of tropical storms Ernesto and Florence during the experiment in the summer of 2006 provided very high sea states. Estimates of mode wander and spread before, during, and after passing of the tropical storms will be presented and compared to existing theories. [Work supported by Office of Naval Research.]

Session 2aAOb

Acoustical Oceanography: Geoacoustic Inversion

N. Ross Chapman, Chair

Univ. of Victoria, School of Earth and Ocean Sciences, P. O. Box 3055, Victoria, BC V8W 3P6, Canada

Contributed Papers

9:45

2aAOb1. Bayesian geoacoustic inversion in a range-dependent environment. Ross Chapman, Yongmin Jiang (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada), and Bill Hodgkiss (UCSD, La Jolla, CA)

In August 2006 a series of experiments was carried out on the New Jersey continental shelf to investigate geoacoustic inversion in a range-dependent shallow water environment. The experimental site was instrumented with multiple oceanographic sensing systems to provide spatial and temporal ground truth of the ocean environment throughout the experiments. The acoustic experiments used multiple receiving systems and sound sources operating over a large frequency band from 50 to 20 kHz. This paper presents results from inversion of low-frequency (<1000 Hz) data obtained along a radial track from one of the 16-element vertical line arrays. The track was surveyed with a chirp sonar to establish bathymetric and subbottom ground truth. Continuous wave tones were recorded at ranges of 1, 3, and 5 km where the source ship held station for several minutes. Bayesian matched field inversion was applied at each of the sites. The effects of unknown range dependence were taken into account by estimating the data error covariance from multiple data windows at each range station. The inversions indicate that the long-range experimental geometry provided reliable estimates for the sea floor sediment parameters in the range-dependent environment. [Work supported by ONR Ocean Acoustics Team.]

10:00

2aAOb2. Ambient noise inverted sonar performance sensitivity due to ocean variability. Juan Arvelo (Johns Hopkins Univ., Appl. Phys. Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099), Daniel Rouseff, and Dajun Tang (Univ. of Washington, Seattle, WA 98105)

Undersea noise was collected on a vertical array affected by ocean currents and internal waves. Such fluctuations are known to negatively affect the accuracy of the inferred bottom loss over grazing angles [Arvelo and Prosperetti, J. Acoust. Soc. Am. **120**, 3257 (2006)]. However, bottom loss is just an intermediate parameter to sonar performance predictions. The bottom loss at the dominant incident angles on the ocean floor is of most importance to the prediction of sound transmission. Therefore, the focus of attention should be on the effect of waveguide dynamics on the predicted acoustic propagation using the estimated bottom loss. Simultaneous transmission loss measurements are compared against predictions from several snapshots of collected element-level wind-driven noise to determine the bias and statistical moments associated with the *in situ* sonar performance estimations across frequency and range. In addition, dominant mode rejection will be applied on a subset of the data that is influenced by nearby ship interference to examine its impact on the robustness of this approach to passive environmental assessment. [This effort was conducted under the auspices of the JHU/APL independent research and development program.]

10:15

2aAOb3. Proof that plane wave attenuation at low frequencies is proportional to square of frequency. Allan D. Pierce and William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02115, adp@bu.edu)

Several published and presented papers by Ingenito [1973], Zhou and Zhang [2005], and others have discussed the attenuation of sound in marine sediments at low frequencies, and another is shortly to appear [Holmes, Carey, Dediu, and Siegmann, JASA Express Letters, 2007], the latter reporting a nonlinear frequency dependence as ω^n, where $n = 1.8 \pm 0.2$; earlier literature gives results with non-integer values of n that are between 1 and 2. The present paper argues that the exponent should be exactly 2 in the limit of low frequencies, given that the sound speed is finite. Plane wave propagation in a general medium allows specific modes for which wavenumber k is a given function $k(\omega)$. Attention is given to those propagating modes where the real part of k is proportional to ω at low frequencies, so that the phase velocity approaches a finite value as the frequency goes to zero. For such circumstances, causality considerations require the imaginary part of k to have a low-frequency dependence that is as the square of frequency or weaker.

10:30

2aAOb4. Geoacoustic inversion using combustive sound source signals. Gopu R. Potty, James H. Miller, Colin J. Lazauski (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882), Preston S. Wilson (Univ. of Texas, Austin, TX 78713-8029), James F. Lynch, and Arthur Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA)

In the summer of 2006, the combustive sound source (CSS) was deployed off the coast of New Jersey during the Shallow Water 2006 experiment (SW-06). CSS generates an oscillating bubble through the combustion of a fuel/oxidizer mixture, which in turn yields a low frequency acoustic pulse [Wilson, Ellzey, and Muir, IEEE J. Ocean. Eng. **20** (1995)]. The depth of these shots was 26 m in water depths of the order of 100 m. Source levels were monitored using a hydrophone placed close to the source. CSS data collected downrange on single hydrophone receiving units (SHRU) are presented. Five SHRUs were deployed by Woods Hole Oceanographic Institution (WHOI) in water depths ranging from 65 to 110 m. The CSS data collected indicate modal dispersion and are used for geoacoustic inversions. These inversions are based on matching the observed and modeled group speed dispersions including Airy phase. Historic sediment data from the location are utilized to constrain the inversions. [Work supported by the Office of Naval Research.]

10:45

2aAOb5. Geoacoustic inversion using the data nullspace projection method. Ying-Tsong Lin, Arthur E. Newhall, and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The data nullspace projection method is developed to reduce the effect of water-column fluctuations on geoacoustic inversions. The basis of this method is to project the bottom information contained in acoustic data onto a data nullspace that is insensitive to the variations of water acoustic properties. To capture more bottom information, the data nullspace needs to be enlarged, which can be done by employing the empirical orthonor-

2a TUE. AM

mal function (EOF) analysis to efficiently decompose the fluctuating acoustic properties in the water column to only few dominant EOF modes. A model resolution matrix is derived to display the preservation of bottom information. The trade-off between preserving bottom information and reducing the effect of water-column fluctuations is discussed. The acoustic data collected from the SW06 experiment, which was conducted on the Mid-Atlantic Bight continental shelf off New Jersey in the summer of 2006, are used to invert for the bottom acoustic properties in the surveyed area. Specifically, the oceanography measurements during the experimental period will provide us with accurate EOF modes for the data nullspace projection method. The efficiency of this projection method is compared with another method that simultaneously solves for bottom and water column properties.

11:00

2aAOb6. A time-frequency distribution-based approach for modal eigenvalue estimation in range-dependent shallow-water waveguides. Megan Ballard (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16802, msd200@psu.edu) and Kyle Becker (Penn State Univ., State College, PA 16803-0040)

In shallow water, geoacoustic properties of the seafloor can be inferred from estimates of modal eigenvalues. For range-dependent environments, horizontal wave numbers associated with propagating modes are sensitive to both changes in bathymetry and sediment properties. Assuming slow variations in the environment, range-dependent wave number estimates can be obtained using short-time Fourier transform (STFT) based approaches. Alternatively, range-dependent wave number estimation can be achieved using time-frequency distributions (TFDs). Traditionally, TFDs have been used for the analysis of signals containing multiple time-varying frequencies. In this work, these methods are applied to horizontal line array data in the spatial domain to reveal instantaneous changes of discrete horizontal wave numbers. In contrast to STFT-based approaches, which assume the environment be comprised of multiple, nearly range-independent segments, the time-frequency distribution approach allows for continuous variations of waveguide properties. Instantaneous wave numbers estimated by TFDs clearly reveal range dependence of the geoacoustic environment through a simple graphic representation. The discrete Wigner-Ville distribution was used for this work with smoothing employed to suppress artifacts caused by closely spaced wave numbers. The utility of using TFDs is demonstrated by application to several range-dependent environments. [Work supported by a NDSEG fellowship and ONR.]

11:15

2aAOb7. Ocean acoustic sensitivity to range-dependent environmental variability. Stan E. Dosso, Michael G. Morley (School of Earth and Ocean Sci., Univ. of Victoria, Victoria BC, Canada V8W 3P6), Diana F. McCammon (McCammon Acoust. Consulting, Waterville NS, Canada), Peter M. Giles, Gary H. Brooke (General Dynamics Canada, Ottawa ON, Canada), Sean Pecknold, and Paul C. Hines (Defence Res. and Development Canada-Atlantic, Dartmouth, NS, Canada)

This paper considers the sensitivity of ocean acoustic propagation data required for sonar performance prediction to range-dependent variability in physical parameters of the ocean environment. In general, the effect on acoustic fields of perturbations to environmental parameters can consist of

both a variation in the field at a point and a spatial shift of the field structure. Accounting for the field shift (via a correlation-based two-dimensional search) can provide a more meaningful measure of sensitivity for many applications. The sensitivity of acoustic fields to the state of knowledge of the bathymetry and of range-dependent variability in seabed geoacoustic parameters is examined both with and without field-shift compensation.

11:30

2aAOb8. Bayesian full-wavefield reflection coefficient inversion and uncertainty estimation. Jan Dettmer, Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC, Canada V8W 3P6), and Charles W. Holland (Penn State Univ., State College, PA 16804)

This paper presents a Bayesian inversion technique to recover seabed geoacoustic properties of multi-layered media from measured broadband reflection-loss data taking full-wavefield effects into account. Seismo-acoustic traces are time windowed into packets with reflection effects of different sets of layers to increasing depth. Each packet is processed to yield reflection coefficients as a function of angle and frequency. The reflection coefficient data are inverted using a full-wavefield forward model. The posterior probability density is sampled with a Gibbs sampler that steps through the packages, using the results for one package as prior knowledge to constrain subsequent inversions. Prior knowledge about the number of layers and their thicknesses is obtained through a separate Bayesian inversion of picked reflection travel times. The posterior probability density for the final package is considered the full solution to the inverse problem and is interpreted in terms of parameter estimates, marginal probability distributions, credibility intervals, and parameter correlations.

11:45

2aAOb9. Uncertainty estimation in Bayesian seismo-acoustic reflection travel-time inversion. Jan Dettmer, Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC, Canada, V8W 3P6), and Charles W. Holland (Penn State Univ., State College, PA 16804)

This paper presents a nonlinear Bayesian inversion for high-resolution seabed reflection travel time data, including rigorous uncertainty estimation and examination of statistical assumptions. Travel time data are picked on seismo-acoustic traces and inverted for a layered sediment sound-velocity model. Particular attention is paid to picking errors that are often biased, correlated, and nonstationary. Non-Toeplitz data covariance matrices are estimated and included in the inversion along with unknown travel time offset (bias) parameters to account for these errors. Simulated experiments show that neglecting error covariances and biases can cause misleading inversion results with unrealistically high confidence. The inversion samples the posterior probability density and provides a solution in terms of one- and two-dimensional marginal probability densities, correlations, and credibility intervals. Statistical assumptions are examined through the data residuals with rigorous statistical tests. The method is applied to shallow-water data collected on the Malta Plateau during the SCARAB98 experiment.

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration: Modeling of Acoustic Cavitation *In-Vivo*

Sheryl M. Gracewski, Cochair
Univ. of Rochester, Mechanical Engineering Dept., 235 Hopeman, Rochester, NY 14627

Diane Dalecki, Cochair
Univ. of Rochester, Mechanical Engineering Dept., Hopeman Bldg., Rochester, NY 14627

Chair's Introduction—8:30

Invited Papers

8:35

2aBB1. Bubble nucleation and cavitation in soft tissue: The initial moments. Charles C. Church (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677) and Vera A. Khokhlova (Moscow State Univ., Moscow 119992, Russia)

In the absence of underlying pathology, biological tissue is expected to be free of gas bubbles. In order for cavitation to occur, the necessary bubbles must be nucleated by the acoustic field. Tiny vapor-filled bubbles continuously form and dissolve in regions of very low interfacial tension. In the presence of an acoustic wave, such microbubbles will act as cavitation nuclei. The probability of nucleation increases with both temperature and rarefactional pressure. Nonlinear propagation and diffraction of high-intensity focused ultrasound (HIFU) in tissue results in the formation of shocks and asymmetric waveforms with a lower peak but longer negative pressure phase than linear analysis predicts; the presence of shocks also greatly enhances energy deposition with a correspondingly more rapid temperature rise in the focal region. The effects of these processes on nucleation rate will be assessed. Post-nucleation bubble motion will then be explored using a modified version of the Keller-Miksis equation for nonlinear bubble dynamics combined with the linear Voigt model for viscoelastic media and a three-component version of the model to simulate a layer of degraded tissue between the gas bubble and the surrounding tissue. Various cavitational phenomena will be discussed. [Work supported by NSBRI SMS00402, RFBR, and DAMD17-02-2-0014.]

8:55

2aBB2. Acoustic bubble simulations for biomedical applications using boundary element method (BEM). Boo C. Khoo (Mech. Eng. Dept., Natl. Univ. of Singapore, Kent Ridge, Singapore 119260), Siew W. Fong, Evert Klaseboer, and Cary Turangan (Inst. of High Performance Computing, Singapore Sci. Park II, Singapore 117528)

High intensity soundwaves, such as shockwaves and HIFU, are widely used in biomedical applications, for example, extracorporeal shockwaves lithotripsy (ESWL) and HIFU prostate cancer treatment. The high pressure in the tissue and nearby fluid often causes formation of cavitation bubbles. Our previous simulation results via BEM approach show complex bubble ultrasound interactions: in certain cases a jet is formed directed away from the nearby tissue while in others, towards it [Fong *et al.*, Ultrasound Biol. Med., **32**(6), 925–942, 2006]. In the present work, pulsed ultrasound microbubble interaction is studied using the same code to provide further understanding to the jetting behavior of the bubbles near (different) biomaterial surface which is critical to the success of tissue ablation and the minimization of auxiliary damages. Separately, we have also simulated shockwave bubble interaction using the BEM code with excellent concurrence to other compressible numerical schemes such as arbitrary Lagrangian-Eulerian and free Lagrange methods [Klaseboer *et al.*, Comput. Methods Appl. Mech. Eng. **195**, 4287–4302 (2006)] and experiments. It is suggested that the (primary) dynamic response of the bubble to soundwaves is mainly inertia controlled since compressibility of the fluid is not modeled.

9:15

2aBB3. The natural frequency of nonlinear oscillation of ultrasound contrast agents in microvessels. Shengping Qin and Katherine W. Ferrara (Dept. of Biomed. Eng., Univ. of California, One Shields Ave., Davis, CA 95616)

Prediction of natural frequency of oscillation of ultrasound contrast agents (UCAs) in microvessels plays an important role in UCA imaging and drug delivery. Here, the natural frequency of a microbubble in microvessels was studied in detail. Three models for blood vessel compliance have been considered, i.e., a rigid vessel, normal compliant vessel, and a stiff vessel which may correspond to the tumor-developing vasculature. Numerical results demonstrate that the natural frequency of a bubble in a rigid vessel is substantially decreased as compared to that in an unbounded field. However, within a compliant vessel, the natural frequency of bubble oscillation increases with decreasing vessel size. When a bubble with a radius of 4 m is confined in a compliant vessel (inner radius 5 m and length 100 m), the natural frequency of bubble oscillation increases by a factor of 1.7 as compared with that in an unbounded field. The natural frequency of bubble oscillation in a compliant vessel decreases with increasing values of vessel rigidity. This model suggests that contrast agent size, blood vessel size distribution, and the type of vasculature should be comprehensively considered for choosing the transmitted frequency in UCA imaging and drug delivery. [Work supported by NIH CA 103828.]

2a TUE. AM

2aBB4. Simulation of an acoustically excited bubble within a compliant vessel. Sheryl M. Gracewski, Hongyu Miao (Dept. of Mech. Eng. and the Rochester Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627), and Diane Dalecki (Univ. of Rochester, Rochester, NY 14627)

Various independent investigations indicate that the presence of microbubbles within blood vessels may increase the likelihood of ultrasound-induced hemorrhage and endothelial cell damage. To explore potential damage mechanisms, an axisymmetric coupled finite element and boundary element code was developed to simulate the response of an acoustically excited bubble centered within a compliant tube. Results will be presented to show the influence of acoustic excitation parameters, tube dimensions, and material properties on bubble response, tube dilation, and stresses induced within the tube wall. Tube dilation occurs during the tensile portion of the acoustic excitation, when the external pressure is lower than the internal pressure. The hoop stress is the maximum induced principal stress. Its peak value occurs as the bubble is expanding, well before the bubble reaches its maximum radius. As the tube thickness, tube radius, and acoustic frequency decreases, the maximum hoop stress increases, indicating a higher potential for tube rupture and hemorrhage. [Work supported by NIH research grant R01HL69824.]

9:55–10:10   Break

2aBB5. Bubble and particle interactions in constrained media. Jianying Cui, Mark F. Hamilton, Todd A. Hay, Yurii A. Ilinskii, Preston S. Wilson, and Evgenia A. Zabolotskaya (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Shock wave lithotripsy and high intensity focused ultrasound produce a variety of cavitation bubble interactions in confined spaces. Densely packed bubble clusters are generated that exhibit strong collective interactions. In lithotripsy, the bubble clusters themselves interact with the surface of a kidney stone. Comminution of the stone generates an accumulation of small particles that are interspersed with, and dynamically coupled to, the cavitation bubbles. Within tissues surrounding the stone, cavitation activity can occur in urinary ducts and blood vessels, constrained by channel walls. We are developing a systematic approach to modeling coupled bubble dynamics in clusters, amidst particles, near surfaces, and in tubes. The model is based on Lagrangian and Hamiltonian formulations of the dynamical equations for interacting bubbles. The initial formulations model the acoustically driven pulsation and translation of clustered spherical bubbles in a free field [Ilinskii *et al.*, J. Acoust. Soc. Am. **121**, 786–795 (2007)]. This model has been augmented to include interaction with spherical elastic particles. Interaction with rigid surfaces, including confinement by tubes with triangular, square, and hexagonal cross sections, has been modeled via the method of images. The relevant model equations will be presented and discussed, and representative simulations will be shown.

Contributed Papers

2aBB6. Biological cell and bubble interactions and behavior in acoustic fields. John S. Allen III (Dept. of Mech. Eng., Univ. of Hawaii-Manoa, Honolulu, HI 96822 alleniii@hawaii.edu) and Pavel Zinin (Hawaii Inst. of Geophys. and Planetology, Univ. of Hawaii-Manoa, Honolulu, HI 96822)

For developing biomedical ultrasound applications such as drug delivery, a more comprehensive understanding of the interaction of an individual cell in an acoustic field with a cavitation bubble at close vicinity is needed. The underlying mechanisms of these phenomenons are not well understood since limited analytical theory exists on the governing mechanics of these interactions. A recent model [Zinin, Allen and Levin, Phys. Rev. E, 2005] for the natural oscillations of a biological cell in a high frequency acoustic field is extended to account for the cells response and deformation from a nearby oscillating bubble. The elastic properties of the cells membrane and viscous properties of the cytoplasma (internal fluid) and the surrounding fluid media are considered in this model. The cells mechanical response is dominated by the surface area modulus (dipole) or shear modulus (quadrupole). The induced stress on the cells membrane from a bubbles oscillation can be calculated and related to membrane transport behavior. For a preliminary study, the influence of linear radial bubble oscillations are examined in detail and potential extensions to the nonlinear regime are discussed. Also highlighted are the potential implications for applications involving bacteria and cancer cells.

2aBB7. Weakly nonlinear bubble pulsations in rigid tubes. Jianying Cui, Mark F. Hamilton, Preston S. Wilson, and Evgenia A. Zabolotskaya (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

In a previous investigation [Cui *et al.*, Proc. 17th ISNA, edited by Atchley *et al.* (AIP, New York, 2006), pp. 229–232], an analytical model was developed for the weakly nonlinear pulsation of a spherical bubble in a channel formed by two rigid parallel plates. The boundary conditions on the plates were satisfied by introducing an infinite line array of image

bubbles. Here we extend this approach to rigid tubes having triangular, square, and hexagonal cross sections. The method of images can also be used for these three geometries because their cross sections tessellate two-dimensional space. Whereas the signals from the images can be summed analytically for a channel formed by parallel plates, numerical summation is required for the three tubes. Resonance curves for an acoustically driven bubble are presented for the response at the fundamental and second-harmonic frequencies as functions of effective tube diameter. The main result is that the resonance curves for all three cross-sectional geometries agree when normalized by the cross-sectional areas. Except when the bubble is within several bubble radii of a tube wall, the results are also nearly independent of the position of the bubble in the tube. [Work supported by ONR.]

2aBB8. Generation of acoustic waves in the eye via cavitation during photoablation of corneal tissue. Mardi C. Hastings (Appl. Res. Lab, Penn State Univ., State College, PA 16804)

During LASIK and other refractive correction surgeries, the shape of the cornea is modified by removal of tissue with high energy laser pulses. The laser beam deposits enough energy over a small spot size (approximately 1 mm diameter) to create a cavitation event from localized heating. Multiple pulses are delivered at different locations on the cornea to remove tissue selectively and reshape its surface in effort to minimize optical wavefront aberrations. Cavitation with each pulse can generate an acoustic wave that propagates within the eye and potentially damage sensory cells in the retina. A mathematical model was created to investigate this phenomenon and understand the effects of laser spot size and pulse repetition frequency on characteristics of the acoustic wave. Motion associated with the cavitation event was treated as a point source on the surface of a spherically shaped eye globe. A spatial map of internal acoustic pressure amplitudes was then created; results indicate that operational parameters can be adjusted to help prevent deleritous effects of cavitation-induced stress waves on the retina during refractive correction surgery.

Session 2aEA

Engineering Acoustics: Underwater Transducers and Arrays

Stephen C. Thompson, Chair

Pennsylvania State Univ., Applied Research Lab., P. O. Box 30, State College, PA 16804

Chair's Introduction—9:00

Contributed Papers

9:05

2aEA1. High drive and duty cycle characterization of single crystal transducers. Harold Robinson, Robert Janus (NUWC Div. Newport Code 1512, Bldg. 1170, 1176 Howell St., Newport, RI 02841), James Powers, and P. David Baird (Progeny Systems Corp., Salt Lake City, UT 84111-5201)

Previous device demonstrations have clearly illustrated how the larger strain energy density and electromechanical coupling factors of single crystal could be used to increase both the source level and bandwidth of Naval sonar projectors. This paper will report on two experiments that conclusively demonstrate that single crystal transducers can withstand high drive levels, elevated duty cycles and extended lifetimes without a significant degradation of performance. In the first experiment, a twelve-element array of PMN-PT tonpilz transducers was driven at its resonance frequency to achieve an acoustic intensity of 5 W/sq. cm. Over a range of duty cycles up to 40% and operating times of up to 30 minutes, the impedance and acoustic output of this array remained completely stable once equilibrium had been reached. In a second experiment, a single crystal cylinder transducer was driven at its resonance frequency for ten minutes at duty cycles ranging from 10% to nearly CW. A temperature rise of only 10C over the duration of the test at the highest duty cycles was measured, with only a 20% increase in admittance and no change in tuned power factor. Both transducers returned to their original state following these tests. [This work is sponsored by ONR.]

9:20

2aEA2. Remote underwater intense laser acoustic source. Ted G. Jones, A. C. Ting, J. Penano, P. Sprangle (Naval Res. Lab., Plasma Phys. Div., 4555 Overlook Ave. SW, Washington, DC 20375), and L. D. Bibee (Naval Res. Lab., SSC, Stennis, MS 39529)

A novel underwater acoustic source is under development in which a tailored intense broadband laser pulse propagates many meters underwater and compresses at a predetermined remote location. Controlled compression of these optical pulses is governed by a combination of optical group velocity dispersion (GVD) and nonlinear self-focusing, resulting in photoionization, localized heating, and shock generation. Configurations involving both atmospheric and underwater laser propagation are also under investigation for applications where airborne lasers generate underwater acoustic signals. The ability to project an acoustic source far from a laser driver, and the increased photoacoustic energy conversion efficiency from the photoionization-shock generation mechanism, are significant improvements on previous laser acoustic source development. Recent experiments included the first demonstration of underwater acoustic generation using an intense lens-focused broadband laser pulse. Initial measurements of sound pressure levels yielded 170 dB for an optical pulse energy of 3 mJ, and are in the range of useful levels for some Navy applications. Optical GVD was precisely measured. Further acoustic measurements of this source, including photoacoustic energy conversion efficiency, acoustic power spectrum, and radiation pattern are underway. Experimental results will be presented.

9:35

2aEA3. Dynamics of nonlinear piston-generated and vertical-array-generated foci. B. Edward McDonald (Naval Res. Lab., Code 7145, Washington, DC 20375), Shane C. Walker, and William A. Kuperman (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

Recent work using a vertical array of linear sources to produce a distant virtual source array [Walker *et al.*, J. Acoust. Soc. Am. **120**, 3007 (2006)] motivates investigation of the possibility of generating a nonlinear virtual source array. First, the nonlinear focal properties of a single piston as a function of amplitude and piston aperture are examined using the NPE [McDonald and Kuperman, J. Acoust. Soc. Am. **81**, 1406 (1987)] and compared with results from the KZK equation [J. Acoust. Soc. Am. **119**, 3618 (2006)]. Next, results are shown for a phased vertical array of spherical explosive sources, which produce Mach stemlike structures as the nonlinear component waves interact. Finally, the fully three-dimensional problem of multiple piston beams focusing in a predetermined pattern is described. [Work supported by the Office of Naval Research.]

9:50

2aEA4. Grating lobe reduction in transducer arrays through structural filtering of supercritical plates. Brian E. Anderson and W. Jack Hughes (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804)

The effect of placing a structural acoustic filter between water and the transducer elements of an array is investigated to help reduce undesirable grating lobes. A plate is mounted to transducer elements with a thin decoupling polyurethane layer between the transducers and the plate. The plate acts as a radiation/incidence angle filter to pass energy at angles near normal incidence, but suppress energy at large incidence angles. Grating lobe reduction is achieved at the expense of limiting the available steering of the main lobe. Within this steer angle limitation, the main lobe can be steered as normal while the grating lobe level is reduced by the plates angular filtering. The insertion of a plate structural filter provides an inexpensive and easily implemented approach to extend usable frequency bandwidth with reduced level grating lobes, without increasing the number of array elements. Even though some data matches theory well, a practical material has yet to be found that possesses optimal material properties. To the authors' knowledge, this work represents the first attempt to advantageously utilize a plate to provide angular dependent sound transmission filtering above the plates critical frequency (the supercritical frequency region). [Work sponsored by ONR Code 333, Dr. David Drumheller.]

10:05

2aEA5. Analysis of a second order endfire line array using SYSNOISE. Thomas Burns (Starkey Labs, Inc., 6600 Washington Ave. S., Eden Prairie, MN 55344)

It can be shown canonically that two equal-amplitude dipoles configured linearly can be summed to yield a directivity index of 9 dB if the directional axes are pointed in 180 deg opposition, and the rear-facing dipole is delayed by the appropriate amount of time. For a 9 mm end-to-

2a TUE. AM

end aperture, the rear facing dipole should be delayed by 7.25 microsec before it is summed with the front-facing dipole in order to achieve the 9 dB. The purpose of this study is to compare the canonical results to those of SYSNOISE. The two dipoles are modeled within SYSNOISE as: 1) four monopoles, 2) four meshed spheres with ka1, and 3) a 9 mm by 1 mm meshed sheet of zero thickness. For each of these three cases, the amplitude and phase of the two dipoles is summed as described above, and the directivity pattern and directivity index is computed by integrating the responses from 614 field points located at equal 10 deg azimuth and elevation angles at 1 m radius from the line array origin. Although the four monopoles and four meshed spheres yielded excellent results at 1 kHz and 6.3 kHz, the results from the meshed sheet was degraded. Further analysis indicated that the phase computation from SYSNOISE was the culprit.

10:20

2aEA6. Physical models of transducer coupling in dense arrays. Kassiani Kotsidou and Charles Thompson (Elec. and Comput. Eng. Dept, Univ. of Massachusetts, Lowell, 1 Univ. Ave., Lowell, MA 01854)

In this work we will examine the coupling between direct sound radiators. It will be shown that at low frequencies, the mutual impedance between a pair of transducers is analogous to that of two simple sources. Hence the magnitude of their interaction will be inversely proportional to their separation distance. The phase is proportional to the propagation time between a transducer pair. The implication of this result is explored in the design of a densely packed arrays, will be examined.

10:35

2aEA7. Fourier-operator methods for parametric array analysis: Theory versus Experiment. Harvey Woodsum and William Hogan (Sonetech Corp., 10 Commerce Park North, Unit 1, Bedford, NH 03110)

Fourier operator methods pertaining to nonlinear scattering of sound by sound, previously developed by the present authors [J. Acoust. Soc. Am. **95**(5), part 2 (2PA14)], have been refined into computer codes, to allow prediction of parametric array performance over a wide range of circumstances, to include parametric array far field and near field, with linear medium attenuation as well as saturation effects fully accounted for. Experimental data sets have been obtained in the course of parametric source calibrations, which allow precise comparison with the new model, showing favorable comparisons in all cases.

10:50

2aEA8. A theory of low frequency parametric arrays in shallow water. Roger Waxler and Thomas G. Muir (Natl. Ctr. for Physical Acoust., University, MS 38655)

Consider a highly directional transducer radiating horizontally in shallow water at two distinct frequencies. The difference frequency beam created by the nonlinear response of the medium is studied. In the far field, the difference frequency beam may be expressed as the product of a horizontal directivity function with a pressure field depending only on depth and range. The horizontal directivity function is identical to the directivity obtained by restricting the classical theory of parametric arrays in free space to two spatial dimensions. The range and depth dependence, however, is that which would be produced by the original transducer radiating at the difference frequency alone, times a small constant.

TUESDAY MORNING, 5 JUNE 2007   ALPINE WEST, 10:00 A.M. TO 12:00 NOON

Session 2aED

Education in Acoustics: Hands-On Demonstrations for Elementary School Students

Uwe J. Hansen, Chair

64 Heritage Dr., Terre Haute, IN 47803

Students and scientists will staff 20 experiment stations for local upper level elementary students who will perform the experiments. These students will thus be exposed to a variety of acoustics principles and get some practical hands-on experience with a number of research tools. Regular ASA participants are welcome to the session, as long as their participation does not interfere with student hands-on activities.

Session 2aMU

Musical Acoustics: General Topics in Musical Acoustics

George A. Bissinger, Chair

East Carolina Univ., Dept. of Physics, Greenville, NC 27858

Chair's Introduction—9:00

Contributed Papers

9:05

2aMU1. Three-dimensional normal mode vibration plus acoustic analysis of two exemplary old Italian violins. George Bissinger and Danial Rowe (Phys. Dept., East Carolina Univ., Greenville, NC 27858)

A Polytec three-dimensional scanning laser system capable of measuring in-plane and out-of-plane violin corpus vibrations was used to measure mobilities in the top, ribs, and back of two exemplary old Italian violins, the 1715 Titian Stradivarius and the 1735 Plowden Guarnerius del Gesu, plus partial scans of the Willemotte Stradivarius (1734) and a Joseph Curtin (2006), back plate and top plate, respectively. The ratio of in-plane to out-of-plane mobility was extracted for the first time for these violins over the frequency range of 0–5 kHz. Comparing only the averaged top plate mobilities of the 1715 Stradivarius and the 1735 Guarneri del Gesu with the 2006 Joseph Curtin revealed no important old-new trend. Among the low-lying (below 600 Hz) signature corpus modes, fit results for total damping revealed higher old Italian total damping only for the upper first corpus bending mode. Analysis of higher frequency damping behavior is currently under way. These measurements were accompanied by acoustic measurements over a sphere in an anechoic chamber of all these violins (plus a 2006 Sam Zygmuntowics violin) to compute radiation efficiencies for each violin.

9:20

2aMU2. The acoustic intensity and directivity of the modern concert harp was investigated. Elizabeth Panzer and Jonas Braasch (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

The acoustic intensity and directivity of the modern concert harp was investigated in two distinct scenarios. Previous studies [Bell and Firth, Acustica (1989); Gautier and Dauchez, Appl. Acoust. (2004)] have relied on mechanical excitation of the strings. Their findings identify two main sources for acoustic radiation, the soundboard and the soundbox holes, the second of which proved to be the dominant source in three frequency ranges. To gain better knowledge of the acoustics of the harp under natural playing conditions, new measurements were conducted in which harp players excited the instrument. One particular goal was to demonstrate the effect of the player on the sound radiation. The absorption and diffraction introduced by the performer diminished the relative importance of the radiation from the soundbox holes. A detailed description of the acoustic intensity field with and without the player is given. In addition, the acoustic profiles of a variety of extended techniques used in contemporary music were investigated. A complete timbral description of these techniques demonstrates the breadth of character the modern harp can produce.

9:35

2aMU3. Acoustic coupling of drum-set cymbals. Edmund Spencer and Paul Wheeler (Elec. and Computer Eng. Dept., Utah State Univ., 4170 Old Main Hill, Logan, UT 84322)

When playing the standard drum set there is a definite audible coupling between the crash cymbal and the splash cymbal. The purpose of this paper was to evaluate this coupling by studying the spectral signature of each cymbal. Separate recordings were made of a 16-in. (40-cm) thin dark crash cymbal and a 10-in. (25-cm) splash cymbal in an anechoic chamber. Spectra Plus software was used to extract the spectral signature of each cymbal. The two cymbals were then placed in proximity with each other and new recordings were made, striking each cymbal in turn. New combined spectral signatures were extracted showing a change in signature for each cymbal due to cross coupling with the other cymbal. The signatures of sympathetic vibrations produced by the unexcited cymbals were then extracted from the combined spectral signatures using MATLAB software. This experiment provided a better understanding of the nature of the coupling between cymbals and provided the needed data for future study of the actual coupling mechanism.

9:50

2aMU4. Application of cochlear analysis techniques to finding percussive events in electro-acoustic music. John Anderson Mills, III (Univ. of Texas, 702 W 32nd St, Austin, TX 78705, nodog@mail.utexas.edu)

Electro-acoustic music is a style of music for which often the pressure signal itself is the only consistent, objective representation of the music. A project has been undertaken at The University of Texas to automatically extract audio properties from recordings of electro-acoustic music to create a new, visual representation, and the first property chosen to extract is a measure of the "single damped percussive events" (SDPEs) found in a piece of music. In order to accurately measure the performance and motivate the development of an SDPE finding algorithm, 26 human participants ranked percussive and near percussive sounds according to how much the sounds were like an SDPE and marked at what point the sounds were no longer an SDPE. The results of this data collection show that rise time is the most important percussive cue and that tonal content, spectral filtering, and ring modulation have little effect on judging a sound to be an SDPE. The preliminary algorithm has been tested against the human performance and the results are presented here. The ultimate goal of the algorithm is to produce a "percussivity profile" of a piece of electro-acoustic music.

10:05

2aMU5. Sinusoidal wavelet transforms for pitch estimation of music signals. Charu Pathak, V. G. Das, V. Prem Kumari, V. Prem Pyra (Dept. of ECE, Jaypee Inst. of Eng. and Technol., NH-3, Agra Bombay Hwy., Raghogarh, Guna (M.P.), India), V. G. Das, V. Prem Kumari, and V. Prem Pyara (Deemed Univ., Agra, U.P. India)

Pitch is related to fundamental frequency in an audio signal. It is a perceptual measure and depends on the auditory system. Discrete Fourier transforms which are most commonly used provide same resolution through the entire range of frequencies and give the output as an average over the entire bin. Resolution of human auditory system varies logarithmically which maps just as the multiresolution analysis based on wavelet transforms. Sinusoidal wavelet transforms are proposed here to detect two continuously occurring signals to a good resolution both in frequency and time. The accuracy can be increased to a desired value by increasing the number of scaling factors used. The technique can be used for note identification and musical instrument recognition through pattern classification.

3061   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3061**

2a TUE. AM

Session 2aNS

Noise: Workshop on Standardization for Soundscape Techniques I

Brigitte Schulte-Fortkamp, Cochair

Technical University Berlin, Inst. of Fluid Mechanics and Engineering Acoustics, Einsteinufer 25, D-10587 Berlin, Germany

Bennett M. Brooks, Cochair

Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, CT 06066

Invited Papers

8:30

2aNS1. Introduction to workshop goals.　Brigitte Schulte-Fortkamp　(Tech. Univ. Berlin, 10587 Berlin, Germany) and Bennett M. Brooks　(Brooks Acoust. Corp., Vernon, CT 06066)

The goal of this workshop is to further explore standardization as a possible future advancement in the evolving field of soundscape measurement, analysis, and design. The workshop will begin in the morning with an introductory keynote address with all participants to review the topics, objectives, and procedures for the day's discussion. The participants will then break out into small subgroups to define key issues and develop action plans. The entire group will reassemble to report, assess, and prioritize the proposed actions in each area.

8:40

2aNS2. Better soundscapes for all—a worthy goal.　Brigitte Schulte-Fortkamp　(Tech. Univ. Berlin, 10587 Berlin, Germany) and Bennett M. Brooks　(Brooks Acoust. Corp., Vernon, CT 06066)

Creating and maintaining soundscapes, or outdoor acoustical environments, that delight the listener may seem to be a difficult task. The perception of the soundscape can provide comfort, tranquility, and needed information to the person concerned, or may be a source of annoyance. Engineering and aesthetic soundscape design may only proceed based on a complete characterization of the acoustical environment, including the nature of the sound sources and the reactions of the perceivers. A powerful tool in this endeavor is the emerging discipline of soundscaping. Soundscaping provides for the measurement, analysis, and design of environmental sound by applying the knowledge of both science and community experts. Much fundamental and practical research has been conducted to establish the bases for the soundscape field. The next step is for researchers and practitioners to standardize the available soundscape techniques to allow for more direct comparison of test and survey results and wider application in design.

9:00

Working Groups, Part 1

The combination of physical acoustical measurements with scientific evaluation of perceptual responses to environmental sound, known as soundscaping, is an essential method for the assessment and actualization of positive outdoor environments. Working groups will be organized around the following areas of interest, as expressed by persons responding to the workshop announcement or at meeting registration. Working groups will develop presentations on problems, solutions, and recommended actions. To further develop and refine the methods of soundscaping, the following are needed:

(1) Catalog of correlations between physical parameters and perceptual responses.

(2) Standardization of a terminology lexicon of soundscape descriptors.

(3) Standardization of measurement procedures.

(4) Recommendations for perceptual evaluation and analysis.

Each working group discussion will focus on (1) what has been done in the past, (2) what should be done now, and (3) how may this be accomplished?

10:40

Working Groups, Part 2

Working group discussions will continue, focusing on: (1) what has been done in the past, (2) what should be done now, and (3) how may this be accomplished? There may be some realignment of the participants in the groups.

Session 2aPAa

Physical Acoustics: Atmospheric Acoustics

Michael J. White, Chair

U.S. Army ERDC/CERL, P. O. Box 9005, Champaign, IL 61826-9005

Chair's Introduction—8:00

Contributed Papers

8:05

2aPAa1. Surface waves in the nocturnal boundary layer. Carrick L. Talmadge, Roger Waxler, Xiao Di, and Kenneth E. Gilbert (Univ. of Mississippi, Natl. Ctr. for Physical Acoust., Coliseum Dr., University, MS 38677)

Experimental observations are presented for pulse propagation in several different types of nocturnal boundary layers. It is shown that, in all cases, a strong surface wave arrival is present, even when other arrivals are weak or absent. Thus, it is argued that the surface wave arrival is a ubiquitous feature of sound propagation in the nocturnal boundary layer. Theoretical reasons are given to support the validity of the argument. Some implications for practical applications are discussed. [Research supported by the U. S. Army TACOM-ARDEC at Picatinny Arsenal, New Jersey.]

8:20

2aPAa2. The influence of angle of incidence at a forest edge on acoustic propagation. Michelle E. Swearingen, Michael J. White, Patrick Guertin, Jeffery Mifflin (Eng. Res. and Development Ctr., Construction Eng. Res. Lab., P.O. Box 9005, Champaign, IL 61826, michelle.e.swearingen@erdc.usace.army.mil), Donald G. Albert, Stephen Decato (Cold Regions Res. and Eng. Lab., Hanover, NH 03755), Arnold Tunick (US Army Res. Lab., Adelphi, MD 20783), and Timothy Onder (Construction Eng. Res. Lab., Champaign, IL 61826)

Acoustic propagation through a forest edge has implications for noise mitigation and acoustic detection systems. While the acoustic significance of this unique environment has been noted in the past, it has not been studied in detail. Acoustic signals that have propagated through a forest edge yield complicated pressure time histories for receivers both within and outside the forest. A recent field experiment was designed to measure acoustic dispersion and attenuation spectra in three distinct regions: open field, forest edge, and deep forest. Ground impedance, meteorological, and biomass structure data were also obtained. This presentation will investigate received signals in the open, edge, and just inside the forest as a function of the angle of incidence between the direct propagation path and the forest edge.

8:35

2aPAa3. Source location using a distributed array of aerostats. Xiao Di, Tau Shi, and Kenneth E. Gilbert (Univ. of Mississippi, Natl. Ctr. for Physical Acoust., Coliseum Dr., University, MS 38677)

An elevated microphone on an aerostat generally can detect a source at a considerably longer range than a microphone on the ground. However, because of the balloon motion, an array of microphones on a single aerostat generally cannot determine a bearing to a source as accurately as a ground-based array. To obtain both long detection ranges and accurate bearings, tests have been done using single microphones on aerostats having a horizontal separation of approximately 100 m. With a long horizontal baseline between the microphones, the bearing estimate is much less sen-

sitive to the balloon motion. Thus, the balloon motion does not have to be precisely monitored. Results from preliminary tests will be presented and discussed. [Research supported by the U. S. Army TACOM-ARDEC at Picatinny Arsenal, New Jersey.]

8:50

2aPAa4. Use of wind and turbulent stability for classifying mean refraction. Michael J. White and Michelle E. Swearingen (US Army ERDC/CERL, P.O. Box 9005, Champaign, IL 61826)

The functional shapes of mean wind and temperature profiles in the boundary layer above the ground depend most strongly on the wind speed and turbulent stability scale size. By use of these two parameters and the wind direction, it is possible to substantially characterize the mean refraction for sound propagation to distances up to about 10 km. This presentation investigates ranges of values and the sensitivity of the vertical effective sound speed profiles to these variations and considers the mean refraction as an objective measure for classifying horizontal propagation.

9:05

2aPAa5. Low-frequency wind noise reduction by spherical windscreens. Richard Raspet, Jeremy Webster, and Jiao Yu (Dept. of Phys. and Astron. and the Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Spherical windscreens are surprisingly effective at reducing wind noise at frequencies for which the turbulence scale is much larger than the windscreen. In the 1930s, Phelps postulated that the low-frequency reduction of windscreens could be explained by assuming that the dc flow pressure distribution around a sphere also applied to low-frequency fluctuations and wind noise. The area average of the dc pressure distribution was then calculated and shown to be smaller than the stagnation pressure. Morgan extended this idea by measuring the pressure distribution around a porous foam windscreen. Recent measurements [Webster *et al.*, J. Acoust. Soc. Am. **118**, 2009 (2005)] have shown that the pressure fluctuations at low frequency do not follow the dc distribution and that the velocity and pressure fluctuation correlations are drastically reduced by the presence of the windscreen. A simple calculation demonstrates that the decorrelation of the pressure fluctuations is sufficient to explain the observed low-frequency reduction afforded by the 9.0 cm diameter windscreen. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

9:20

2aPAa6. An investigation of reverberation in an urban environment. Brad Libbey and James Perea (Army RDECOM CERDEC NVESD, 10221 Burbeck Rd., Fort Belvoir, VA 22060)

Localization of an acoustic blast source in urban battle spaces is frequently complicated by multiple arrivals resulting from reflections within the acoustic environment. Temporal windowing can be used to isolate the direct arrival in situations where a direct path between source and receiver exists. However, if the source is located around a corner from a receiver,

3063   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3063**

2a TUE. AM

the first arrival would lead to an erroneous source location estimate. These complications have prompted the authors to perform initial measurements of reverberation in a semi-urban environment. These measurements will assess the salience of reverberation as a means for estimating range to a blast source.

9:35

2aPAa7. Time-domain boundary conditions in atmospheric acoustics. Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80305 and Dept. of Phys., New Mexico State Univ., Las Cruces, NM 88003), Sandra L. Collier, David H. Marlin (U.S. Army Res. Lab., Adelphi, MD 20783), D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03755), David F. Aldridge, and Neill P. Symons (Sandia Natl. Labs., Albuquerque, NM 87185)

In atmospheric acoustics, formulation of a reactive, time-domain boundary condition (TDBC) at the ground surface is a challenging problem since many commonly used models of the ground are noncausal and have a slowly decaying response. A correctly formulated TDBC can significantly simplify finite-difference time-domain simulation of outdoor sound propagation, which is a rapidly developing field in computational acoustics. In the present paper, approaches are developed that enforce causality and improve computational efficiency of TDBCs. First, an approach is developed that allows one to derive a causal TDBC for any impedance model of the ground. A Pade approximation is used to represent the characteristic admittance. Then, using fractional derivatives, a causal TDBC is formulated. This method is illustrated by formulation of TDBCs for the Zwikker-Kosten (ZK) and Attenborough models of the ground. Furthermore, by approximating the fractional derivatives as a summation of decaying exponentials, an effective recursive algorithm for implementation of these TDBCs is outlined. Second, it is demonstrated that approximating the known TDBC for the ZK model as a summation of decaying exponential functions significantly improves the computational efficiency of the method.

9:50

2aPAa8. Impulse propagation in the nocturnal boundary layer: Experimental observations outdoors with a vertical array. Roger Waxler, Carrick L. Talmadge, Xiao Di, and Kenneth E. Gilbert (Natl. Ctr. for Physical Acoust., University, MS 38655)

As observed experimentally by Chunchuzov et al. [J. Acoust. Soc. Am. **88**, 455 (1990)], impulsive signals propagating in the downward refracting atmospheres typical of the nocturnal boundary layer have a canonical long-range form: they consist of multiple arrivals followed by a low frequency tail. Theoretical arguments that such behavior is canonical and an identification of the low frequency tail with the ground-coupled surface mode have been recently presented [J. Acoust. Soc. Am. **117**, 2532 (2005)]. In November 2006, pulse propagation experiments were performed at night on farmland in the Mississippi River Delta. The source used was a propane cannon and the receiver was a 10 meter tower with 16 equally spaced microphones microphones located 1.6 km downwind from the source. Meteorological data was collected from three sensor stations mounted on the tower, six attached to a tethersonde line and a sodar device. Results from these experiments will be presented and discussed.

10:05–10:20   Break

10:20

2aPAa9. Maximum likelihood estimation of the angle of arrival for waves propagating in atmospheric turbulence. Sandra Collier (U.S. Army Res. Lab., Adelphi, MD 20783), Vladimir Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80305), and Keith Wilson (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03755)

Physical effects such as refraction from temperature and wind gradients and scattering from atmospheric turbulence can have a significant impact on acoustic beamforming. Many current beamformers only account for the medium effects through use of an effective sound speed. Here we develop a maximum-likelihood estimator (MLE) for the angle of arrival of

an acoustic signal that directly accounts for the effects of scattering from atmospheric turbulence. The MLE utilizes a statistical model for the received signal that is based on the theory of wave propagation in a random medium and realistic atmospheric turbulence models. The resulting MLE algorithm is tested using simulated data sets and is found to outperform an estimator that assumes a homogeneous medium.

10:35

2aPAa10. Acoustic propagation in the vicinity of an isolated building. W. C. Kirkpatrick AlbertsII and John M. Noble (U.S. Army Res. Lab., Attn: AMSRD-ARL-CI-ES, 2800 Powder Mill Rd., Adelphi, MD 20783)

The literature regarding urban acoustics has often been concerned with the propagation of sound through street canyons typical of residential areas in large cities, while sparsely built suburban areas have received little attention. As a fundamental case of acoustic propagation in suburban areas, the study of the influence that an isolated building has on propagating sound is a necessity in order to determine the processes (wall impedance, ground impedance, diffraction, diffusion, etc.) required to model the interaction of propagating acoustic waves with an isolated building. As such, an experimental and theoretical investigation of this fundamental case is currently being performed on isolated buildings of construction typical to North American suburban areas. Experimental results, using both continuous and impulsive excitation sources, for propagation near a single story, gabled roof building typical of some suburban areas will be presented.

10:50

2aPAa11. Infrasound propagation in unstable atmospheric layer. Konstantin Naugolnykh (Univ. of Colorado/Zeltech, Boulder, CO 80305) and Samuil Rybak (N. Andreev Acoustics Inst., Moscow, Russia)

Cooling of air in upward convection flow makes water vapor supersaturated. This is a nonequilibrium state of fluid where effect of instability can be developed in the process of infrasound propagation (Nemtsov, 1989; Netreba, 1996). Atmospheric convection flow has a helical character, which introduces an additional source of instability (Moiseev et al., 2000). The basic equation which takes into account both effects is obtained in the present paper on the basis of general consideration of the relaxation process of unstable medium. Solutions of this equation indicate the effect of infrasound wave amplification and its nonlinear evolution. Corresponding analyses of them make it possible to clarify the principally important aspects of cyclone infrasound precursors.

11:05

2aPAa12. Coherence function of a wave propagating through intermittent turbulence. Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80305 and Dept. of Phys., New Mexico State Univ., Las Cruces, NM 88003), D. Keith Wilson (U.S Army Engineer Res. and Development Ctr., Hanover, NH 03755), and George H. Goedecke (New Mexico State Univ., Las Cruces, NM 88003)

An important characteristic of atmospheric turbulence is intermittency, i.e., the tendency of turbulence to occur in bursts of activity. Intermittency can affect the statistical moments of a sound wave propagating in a turbulent atmosphere. To address this problem, in this paper a quasi-wavelet (QW) model of intermittent temperature fluctuations is developed. In the model, QWs become less and less space filling as their sizes decrease. Using this model, formulas for different statistical characteristics (variance, kurtosis, and 3D spectrum) of intermittent temperature fluctuations are derived. Then, the obtained 3D spectrum is used to calculate the coherence function of a sound wave propagating in a turbulent atmosphere. The coherence function is one of the most important statistical moments of a sound field needed in many applications. The effects of intermittency of turbulence on the coherence function are studied. The calculated coherence function is also compared to that obtained by a different approach, in which the structure parameter of refractive index fluctuations is assumed to be a random function along the sound propagation path. The results obtained in this paper can be used in other problems, e.g., to study coherence of seismic waves for line-of-sight propagation through intermittent geological heterogeneities.

Session 2aPAb

Physical Acoustics and Education in Acoustics: Academic Genealogy (Poster Session)

Steven L. Garrett, Chair

Pennsylvania State Univ., Graduate Program in Acoustics, P. O. Box 30, State College, PA 16804-0030

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. To allow for extended viewing time, posters will remain on display until Wednesday at 6:00 p.m.

2aPAb1. One view of the Rudnick-Putterman UCLA-acoustics dynasties. Steven Garrett (Penn State Grad. Progam in Acoust., State College, PA 16802)

During the second half of the 20th Century, the research group headed by Isadore Rudnick augmented then continued to this day by Seth Putterman, brought the formalism and techniques of classical acoustics and phenomenological physics to problems in the quantum mechanics of condensed matter systems. While doing so, Izzy advised 32 Ph.D. students that graduated between 1951 and 1986, half of which went on to academic careers of their own (excluding two of his sons who are also academic scientists), making the UCLA-acoustics family tree extraordinarily dense and complex. Although Putterman's academic roots are traced all the way back to Stefan, Boltzmann, Ehrenfest, and Uhlenbeck, it appears that Rudnick's inspiration was divine in its origin. Traveling forward from my time as a Rudnick-Putterman graduate student in the mid-1970s, several clusters of my academic siblings (fellow graduate students and post-docs) can be identified that exerted a strong influence on acoustics research and education at the Naval Postgraduate School and Penn State, many of whom were influenced by the Swift-Wheatley-Migliori (Thermoacoustics) Group at Los Alamos National Laboratory.

2aPAb2. The degree family tree of Brian E. Anderson. Brian E. Anderson (Los Alamos Natl. Lab., Geophys. Group, M.S. D443, Los Alamos, NM 87545, bea@lanl.gov)

The degree family tree of Brian E. Anderson will be presented. The poster will display evidence of the strong connection between the Graduate Program in Acoustics at The Pennsylvania State University and the Acoustics Research Group of the Department of Physics and Astronomy at Brigham Young University. Currently Brian represents a branch with no fruit, but Brian hopes to someday bear much fruit (advise graduate students as a faculty member).

2aPAb3. Harvard academic tree leading back to Helmholtz. David T. Blackstock (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, and Mech. Eng. Dept., UT Austin, 1 University Station C2200, Austin, TX 78712-0292)

From 1946 to 1971, 30 Harvard Ph.D. degrees were earned in F. V. (Ted) Hunt's Acoustics Research Laboratory. Although Hunt was on my committee, and I'll always consider myself one of "Hunt's boys," William P. Raney was my Ph.D. (Harvard, 1960) supervisor. My direct ancestral line thus begins with Raney, who did his doctoral research at Brown University (Ph.D. 1955) under Peter J. Westervelt. The lineage going further

down the tree is as follows: Westervelt (MIT, 1951), Richard H. Bolt (UCLA, 1939), Vern O. Knudsen (University of Chicago, 1922), Robert A. Millikan (Columbia, 1895), Michael I. Pupin (Berlin, 1887), Hermann von Helmholtz (Berlin, medicine, 1842). Now go up the tree. I had two Ph.D. students at the University of Rochester, Donald B. Cruikshank (1968) and James C. Lockwood (1971). My first Ph.D. graduate from the University of Texas at Austin was F. Michael Pestorius (1973); eight younger siblings followed. A bonus was being stepfather for Mark F. Hamilton (Penn State, 1983) when his Penn State supervisor, Frank Fenlon, died in 1981. I was also blessed with 22 master's students at Texas.

2aPAb4. The academic family tree of Katherine Safford Harris. Fredericka Bell-Berti (Dept. of Speech, Commun. Sci., & Theatre, St. Johns Univ., Queens, NY 11439), Lawrence J. Raphael (Adelphi Univ., Garden City, NY 11530), and Diane Kewley-Port (Indiana Univ., Bloomington, IN 47405)

At Haskins Laboratories Kathy Harris lead a research team of engineers, programmers, technicians, ENT physicians, and fellow scientists to study speech production. As the reputation of this innovative research program grew, students, post-doctoral fellows, and visiting scientists from many universities and international laboratories came to join Kathy in this effort. Her contributions to training doctoral students grew when she became a faculty member at the CUNY Graduate Center. Kathy Harris was a pioneer role model, especially, but not exclusively, for women scientists. Her enthusiasm and dedication to science attracted a host of students and young scientists who have gone on to academic research careers in speech communication. Kathy nurtured her students, always ready to listen and encourage, guiding but never pushing them through their degrees, and always keeping sight of the person.

2aPAb5. F. V. Hunt and acoustics at Harvard University. Wayne M. Wright (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Soon after his arrival from Ohio as a graduate student of G. W. Pierce in 1925, Frederick Vinton Hunt established himself as a Harvard scientist to be considered seriously. He submitted two doctoral theses, joined the faculties of physics and applied physics, and in 1940 graduated his first doctoral student, Leo Beranek. He founded and directed the war-time Harvard Underwater Sound Laboratory for the Navy, and he mentored a total of 70 graduate students and postdoctoral fellows prior to his retirement in 1971, a year before his death. While the acoustics tradition at Harvard effectively ended with Ted Hunt's departure, a number of individuals that he mentored have developed and supported acoustics programs elsewhere in academia. Others have made significant contributions in government

and industry, and several have followed his lead in contributing much to the Acoustical Society of America. Beyond his personal impact, his legacy includes 29 recipients of the Society's F. V. Hunt Postdoctoral Fellowship, the first of which being Steven Garrett (1978–79, now at Penn State), and the recent production under ASA auspices of a CD that makes generally available the 61 Technical Memoranda from Hunt's Acoustics Research Laboratory.

2aPAb6. Jiri Tichy and the Penn State—Brigham Young University connection. Scott D. Sommerfeldt and Timothy W. Leishman (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

Jiri Tichy held the position of Chair of the Graduate Program in Acoustics at Penn State University for many years, and oversaw significant growth in the acoustics program there over those years. During that time, Tichy also had a significant number of graduate students, both at the MS and PhD levels. This poster will attempt to identify most, if not all, of those students, as well as where they went in their careers. In addition, two of Tichy's graduate students both migrated to Brigham Young University (BYU). BYU's acoustics history includes both Carl Eyring and Harvey Fletcher, who both spent some of their careers at BYU. Over the past decade, the BYU acoustics program has enjoyed a resurgence, brought about largely by an influx of Penn State graduates. This has also led to a strengthened connection between Penn State and BYU, and these connections and interrelationships will be outlined in the poster.

2aPAb7. Historical perspective on Robert T. Beyer: Connections and branches surrounding the Brown-Acoustics era. Murray S. Korman (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402) and Stephen V. Letcher (Univ. of Rhode Island, Kingston, RI, 02881-0817)

Starting from a timeline when R.T. Beyer graduated from Hofstra and went to Cornell in the fall of 1942 there are many branches which connect (back in time) to the so-called great-grandfather, grandfather and father advisors leading up to his own research advisor, Dr. Harry Sack, at Cornell. Dr. Sack had been a professor at the Free University of Belgium who worked for Dr. Peter Debye before the war and was a former student. The roots of Debye will be traced backward in time and then Beyer's own family tree of graduate students will be developed. Bruce Lindsay (who worked with Niels Bohr in 1922 in Copenhagen, in his last year of graduate school at MIT) hired Beyer as an instructor starting in the summer of 1945. Professor Beyer's first graduate student was Myron Smith, who used radiation pressure to measure the intensity of sound under his guidance. One of the authors (M.S.K.) was Beyer's last graduate student (1981). The other author (S.V.L.) was a graduate student of Beyer's in the 1960s and co-authored *Physical Ultrasonics* with his mentor. Some threads will be introduced with A.O. Williams and P.J. Westervelt—fellow nonlinear acousticians during Beyer's career at Brown.

2aPAb8. Indiana University in the history of psychological and physiological acoustics. William A. Yost (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sherdian Rd., Chicago, IL 60626, wyost@luc.edu), Constantine Trahiotis (Univ. of Connecticut Health Ctr., Farmington, CT 06032), Charles S. Watson and Donald E. Robinson (Indiana Univ., Bloomington, IN 47405)

The faculty, postdoctoral fellows, and doctoral graduates of Indiana University (IU) in Bloomington, Indiana have played a significant role in the history of the field of psychological and physiological acoustics, especially since 1960. The department of Psychology (now Psychological and Brain Sciences) and that of Speech and Hearing Sciences, along with the former Center for Neural Science, provided a rich environment for many of us to develop as psychological or physiological acousticians. These researchers, their students, and their students students have had a far reaching impact on psychological and physiological acoustics. This poster will describe some of the key IU people, their various contributions, and

the academic connections among these IU researchers and many others in the field. The poster will also discuss IUs success as a major contributor to our field.

2aPAb9. Geneology of Phil Marston. Thomas Matula (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105) and Philip Marston (Washington State Univ., Pullman, WA 99164)

This poster will highlight the geneology of Philip Marston. The poster will follow the geneological progression from Phil's past advisors to his students.

2aPAb10. The acoustics progeny of R. Bruce Lindsay. Peter H. Rogers (School of Mech. Eng., Geogia Inst. of Technol., Atlanta, GA 30332, peter.rogers@gatech.edu)

R. Bruce Lindsay served on the faculties of Yale University from 1923 to 1930, and Brown University from 1931 until his death in 1985, a total of sixty-two years. Over that time he produced at least thirty-five Ph.D. students and a like number of masters students. Many of these students were active in acoustics, held faculty positions, and went on to produce Ph.D. students of their own. This paper attempts to trace the major branches of the acoustics family tree of R. Bruce Lindsay, which is now in at least its fifth generation. The members of the Lindsay family tree have made, and continue to make, significant contributions to acoustics. For example, it is interesting to note that James Finneran, the 2002 winner of the ASAs R. Bruce Lindsay Award, is R. Bruce Lindsays academic great-great grandchild.

2aPAb11. Apfel's apples: A generation of scientists trained at Yale. Jeffrey A. Ketterling (Lizzi Ctr. for Biomed. Eng., Riverside Res. Inst., 156 William St., New York, NY 10038)

Bob Apfel taught at Yale University for over thirty years, advising scores of students during that time. Prior to Yale, Bob trained at Harvard University with Frederick Hunt and the chain of advisors can be loosely traced back, all at Harvard, to 1802. During his time at Yale, Bob graduated 28 Ph.D.'s and trained 12 post-doctoral students in addition to the numerous undergraduates that he advised. Many of the students moved on to become professors themselves and, following the lead of Bob, many are very active in the Acoustical Society of America.

2aPAb12. Over 75 years of acoustics at the National Research Council of Canada. Gilles A. Daigle and Michael R. Stinson (Natl. Res. Council, Ottawa, Canada, K1A 0R6)

The first Director of the Division of Applied Physics brought with him a strong interest in ultrasonics, thereby opening a line of research in acoustics, which has continued unbroken even since. It was George J. Thiessen, Section Head from 1950 to 1975, who rebuilt the acoustics section after the war effort and charted the course the laboratory follows today. It was in the early 50s that Edgar Shaw and Tony Embleton joined the laboratory. They were both students of R. W. B. Stephens, a research supervisor who has taught and inspired a generation of acoustics students and guided them in the attainment of excellence. During the over 50 intervening years, several generations of acousticians have studied and been mentored in the laboratory. Most of the students and post-docs that have spent time in the laboratory now hold faculty, government, or industry positions in acoustics.

2aPAb13. Genealogy of the National Center for Physical Acoustics. Henry E. Bass and James M. Sabatier (NCPA, 1 Coliseum Dr., Univ., MS 38677)

NCPA traces its roots to the Physical Acoustics Research Group which grew from the Department of Physics and Astronomy. The current Department of Physics and Astronomy acoustics effort was initiated with Prof. F. Douglas Shields who received his Ph.D. at Vanderbilt University under the

direction of Prof. Robert Lageman. Professor Shields came to the University of Mississippi in 1959. He directed the dissertation of ten Ph.D. students and seven Masters students. In 1964, Prof. Shields was joined in the department by Prof. Roy Arnold who studied the ultrasonic properties of solids. Professor Arnold directed the dissertation of two Ph.D. students and six Masters students. In 1970, Prof. Randall Peters and Prof. Henry

Bass joined the faculty; Bass doing acoustics in gases with a Ph.D. from Oklahoma State under Tom Winter and Peters doing solid state acoustics. The following year, Ron Carter joined the group doing solid state acoustics. Professor Bass directed the dissertation of nineteen Ph.D. students

and eight Masters students. In 1976, Dr. Hans Bauer, a student of Prof. Kneser, joined the group for one year. Professor Larry Crum, a student of Prof. Stump at Ohio State, joined the faculty in 1978. In 1986, an act of Congress established the National Center for Physical Acoustics and over a period of two years, PARGUM was transferred to NCPA. Professor Ralph Goodman joined NCPA as Director in 1989. Dr. Robert Hickling and Prof. Mack Breazeale joined NCPA in 1988. Professor John Seiner joined NCPA in 1998. A listing of all additional faculty and scientists to join NCPA and the students directed by all will be presented.

TUESDAY MORNING, 5 JUNE 2007　　　　　　　　　　　　　　　　　　　　TOPAZ ROOM, 8:00 TO 11:40 A.M.

Session 2aPP

Psychological and Physiological Acoustics: The Neural Coding of Pitch: Insights from Psychophysics, Neurophysiology, and Brain Imaging

Christopher J. Plack, Cochair
Lancaster Univ., Dept. of Psychology, Lancaster, LA1 4YF, UK

Deborah Hall, Cochair
MRC Inst. of Hearing Research, University Park, Nottingham, NG7 2RD, UK

Chair's Introduction—8:00

Invited Papers

8:05

2aPP1. Pitch perception in normal and impaired hearing. Andrew J. Oxenham (Dept. of Psych., Univ. of Minnesota, Minneapolis, MN 55455)

Despite a century-long debate, there is little consensus on how pitch is coded. This talk will focus on recent psychoacoustic studies that have attempted to shed light on some of the basic principles of pitch coding. The pitch associated with melodies and music is conveyed mainly by lower-numbered harmonics, which are thought to be individually resolved in the peripheral auditory system. When the lower harmonics are removed, the pitch percept becomes weaker and dependent on the phase relations between harmonics, as expected with unresolved harmonics that interact within the passbands of peripheral auditory filters. A number of studies in normal and impaired hearing have shown a connection between harmonic resolvability and pitch strength or salience. However, there are reasons to doubt the causality of this relationship. The evidence for whether resolved harmonics are necessary or sufficient for salient pitch will be reviewed. Finally, some experiments related to the use of pitch cues in segregating speech from competing interferers will be discussed, with particular reference to whether the cues that produce strong pitch salience in psychoacoustic tasks are also those that are most useful in promoting segregation. [Work supported by NIH Grant R01DC05216.]

8:35

2aPP2. Pitch perception in electric and acoustic hearing. Robert P. Carlyon (MRC Cognition & Brain Sci. Unit, 15 Chaucer Rd., Cambridge CB2 7EF, England)

In normal hearing (NH), timing and place-of-excitation cues to pitch are usually inextricably linked. In contrast, by varying the temporal pattern of electrical pulses applied to one channel of a cochlear implant (CI), researchers can investigate the usefulness and limitations of purely temporal cues to pitch. A related technique, which often produces very similar results, is to present NH listeners with trains of acoustic pulses that have been passed through a fixed bandpass filter so as to remove place cues. Furthermore, one can then measure how these temporal patterns are represented in the auditory nerve (AN) by recording, respectively, the electrically and acoustically evoked compound action potential. These physiological measures can then be compared to listeners' pitch judgements to the same stimuli. We have used this approach to show how AN refractory effects can influence pitch, to evaluate competing pitch models, and to show how a code based on first-order intervals can provide a level-independent representation of pitch. We have also examined the possible basis for the relatively poor pitch perception of CI users, particularly at pulse rates above about 300 pps. [Work supported by RNID.]

2aPP3. The neural representation of pitch in the auditory brainstem. Ian Winter (Physiological Lab., Downing St., Cambridge, CB2 3EG, UK, imw1001@cam.ac.uk)

How the pitch of complex sounds is represented in the discharges of single or populations of neurons in the auditory brainstem is still unknown. In the ventral cochlear nucleus (VCN) each of the three main response types, primarylike, chopper, and onset, have all been implicated. Primarylike units maintain a good temporal representation of the pitch, similar to that found in the incoming fibres of the auditory nerve; chopper units can enhance the periodicity of complex sounds when it matches their chopping rate while onset units can show a particularly robust response to the periodicity of complex sound. However, various problems are faced by each of these unit types in explaining all aspects of pitch perception. In the inferior colliculus the theory that has dominated our thinking on the representation of the pitch of complex sounds has been that of the modulation filter bank, which runs orthogonal to the iso-frequency laminae in this nucleus. This modulation filter bank hypothesis has been linked with the periodicity tuning observed in some chopper units in the VCN, but many questions remain to be answered. In this review talk the successes and failures of these ideas will be discussed.

9:35−9:55   Break

9:55

2aPP4. Neural representation of pitch in auditory cortex. Xiaoqin Wang (Dept. of Biomed. Eng., Johns Hopkins Univ., 720 Rutland Ave., Traylor 410, Baltimore, MD 21205)

Pitch perception is crucial for vocal communication, music perception, and auditory object processing in a complex acoustic environment. The neural representation of pitch in the cerebral cortex has long been an outstanding question in auditory neuroscience. Both animal and human studies have implicated the role of the cerebral cortex in underlying pitch perception. Recent findings indicate that there is a specialized region in the auditory cortex of primates that is involved in the representation of pitch. This presentation will review key studies in the identification of a pitch processing center in primate auditory cortex and discuss issues concerning the neural substrate of pitch perception. [Research supported by NIH Grants DC003180 and DC005808.]

10:25

2aPP5. Brain imaging of pitch processing in human auditory cortex. Deborah Hall (MRC Inst. for Hearing Res., University Park, Nottingham, NG7 2RD, UK) and Christopher Plack (Lancaster Univ., Lancaster, LA1 4YF, UK)

Recent data from human fMRI and primate electrophysiological studies suggest that a region near the anterolateral border of primary auditory cortex may be involved in pitch processing. Collectively, these findings present strong support for a single central region that is selective for pitch, irrespective of its spectral characteristics. Most of the human neuroimaging studies addressing this issue have used iterated rippled noise (IRN), a specialized stimulus, to evoke a pitch sensation. In the present fMRI study, IRN did indeed activate the lateral portion of the primary auditory cortex on Heschls gyrus (HG). In contrast, the other pitch-evoking stimuli tested, including a variety of pure and complex tones, and Huggins pitch (a binaural pitch dependent on a combination of information from the two ears), did not activate lateral HG. Although there were individual differences, the most common locus was posterior to HG in planum temporale. We suggest that features of IRN not related to pitch may be responsible for the earlier results, and that it may be premature to assign the pitch center to lateral HG.

Contributed Paper

10:55

2aPP6. Pitch discrimination interference: Monaural and binaural pitches. Hedwig E. Gockel, Robert P. Carlyon (MRC Cognition and Brain Sci. Unit, 15 Chaucer Rd., Cambridge CB2 7EF, UK), and Christopher J. Plack (Lancaster Univ., Lancaster LA1 4YF, UK)

Fundamental frequency (F0) discrimination between two sequentially presented complex (target) tones can be impaired in the presence of an additional complex tone (the interferer) filtered into a remote spectral region [H. Gockel, R. P. Carlyon, and C. J. Plack, J. Acoust. Soc. Am. **116**, 1092–1104 (2004)]. This pitch discrimination interference (PDI) is greatest when the interferer and target have similar F0s. Here, PDI was measured using monaural or diotic complex-tone interferers and Huggins pitch

or diotic complex-tone targets. We first showed that listeners hear a complex Huggins pitch (CHP), approximately corresponding to F0, when multiple phase transitions at harmonics of (but not at) F0 are present. The accuracy of pitch matches to the CHP was similar to that for an equally loud diotic tone complex presented in noise. Similar amounts of PDI were observed for CHP-targets and for loudness-matched diotic complex-tone tone targets. Thus, PDI can occur between stimuli that initially are processed in a different way, making it unlikely that PDI occurs at a very peripheral stage of auditory processing. A corollary is that conventional complex tone and Huggins pitches appear to be processed in common at the stage where PDI occurs. [Work supported by EPSRC Grant EP/D501571/1.]

11:10−11:40
Panel Discussion

Session 2aSA

Structural Acoustics and Vibration and Signal Processing in Acoustics: Sound Source Localization

Sean F. Wu, Chair
Wayne State Univ., Dept. of Mechanical Engineering, 5050 Anthony Wayne Dr., Detroit, MI 48202

Invited Papers

8:30

2aSA1. Nearfield acoustical holography and nonlinear sound fields. J. D. Maynard (The Penn State Univ., University Park, PA 16802)

Nearfield acoustical holography (NAH) has become a well-established method for studying sound radiation from noise sources. One of the advantages of the technique is that it may be implemented with fast computer algorithms. While the customary application assumes linear sound propagation, a current application is for noise from jet engines, where pressure amplitudes in the nearfield are high and sound propagation is nonlinear. A method for treating nonlinear sound fields with NAH, permitting high speed calculation, is being developed. An interesting problem arises from the result that nonlinear effects produce a source field which is distributed in a volume, while NAH treats sound fields on surfaces. A method for converting nonlinear volume sources to surface fields, together with NAH processing, will be presented.

8:55

2aSA2. High-frequency reconstruction of radiation patterns from a rectangular baffled plate using the Helmholtz equation least squares method. Huancai Lu, Sean Wu (Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202), and Donald Keele, Jr. (Harman Intl. Industries, Northridge, CA)

The Audio Engineering Society (AES) Standards require that full-sphere polar measurements of a sound field be taken at a very fine angular increment (1 deg) to ensure the accuracy in describing high-frequency radiation patterns of a source. This super fine angular resolution coupled with a far-field measurement requirement leads to an excessive number of measurements. For example, at 1 m and 1 degree angular resolution we would need more than 64 000 measurement points. To alleviate this difficulty, the HELS (Helmholtz Equation Least Squares)-based near-field acoustical holography is used to reconstruct the radiation patterns. It is shown that HELS can provide accurate reconstruction of radiation patterns for dimensionless frequencies up to $ka = 25$ based on the sound field sampled on a spherical surface at a rate lower than that required by AES Standards. HELS allows for accurate reconstruction even for sample spacing that violates the Nyquist spatial sampling rate in certain directions. In this paper, the convergence of HELS solutions is examined by comparing reconstructed radiation patterns with the analytic solution for a rectangular baffled plate. In particular, the impacts of the number of expansion terms in HELS and measurement points, and that of errors imbedded in the input data on the accuracy of reconstruction are analyzed.

Contributed Papers

9:20

2aSA3. Artefact detection and sound source localization on ships and submarines. Mae L. Seto (Signatures Section, Defence R&D Canada-Atlantic, #9 Grove St., P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

DRDC's Ship/Submarine Signature Management System (SSMS) detects acoustic artefacts based on arrays of sensors strategically mounted at the base of machinery and on the ship/submarine hull. Acoustic artefacts exceed some nominally acceptable level and could compromise a ship/submarine's mission though its impact on the acoustic signature. After artefact detection the next tasks are localization and mitigation. Only localization is discussed. Localization challenges in the ship/submarine environment include machinery being on over very different time scales, multi-path, sound ducting, confusion over multiple sources in a small space, machinery changing speeds, fluctuating broad band levels. etc. SSMS source localization is achieved through judicious grouping of the detected machinery tonals both within and across multiple sensors and their time correlation against machinery activity, This is additionally coupled with knowledge of the machinery and ship layout. Sensors must be synchronously sampled, to within a small tolerance, over an environments that may distribute sensors over a 250 foot length. Such measures

appear to increase the likelihood of a successful source localization to the machinery or structure that created the artefact. The SSMS and its localization algorithm and its validation are reported on.

9:35

2aSA4. Localization accuracy of advanced spatialisation techniques in small concert halls. Enda Bates, Gavin Kearney, Dermot Furlong, and Frank Boland (Trinity College Dublin, Dept. of Electron. & Elec. Eng., Trinity College, Dublin 2, Ireland, batesja@tcd.ie, gpkearney@ee.tcd.ie)

A comparison of several spatialization systems is presented in terms of their localization accuracy under the nonideal listening conditions found in typical small concert halls. Of interest is the effect of real reverberant conditions, noncentral listening positions, and nonsymmetrical speaker arrays on source localization. The data are presented by comparison of empirical binaural measurements and perceptual listening tests carried out using Ambisonics, VBAP, SPAT~, and Delta Stereophony systems. The listening tests are conducted by comparing the localization of phantom sources generated by the spatialization systems to sources generated by reference loudspeakers. The reference and phantom sources are presented at front, side, and back locations about a nine-listener audience, and the systems are tested in a random order with a calibrated 16-loudspeaker array situated nonsymmetrically around the audience area. The binaural

recordings are compared to the subjective measurements of localization accuracy through the interaural level and time difference cues for each source position.

9:50

2aSA5. Localization with vector sensors in inhomogeneous media. Joseph A. Clark and Gerald Tarasek (NSWCCD, Code 7340, 9500 MacArthur Blvd., West Bethesda, MD 20817-5000, joseph.a.clark1@navy.mil)

Vector sensors detect and localize sources of radiated or scattered sound by simultaneously measuring pressure and particle motion near a single point. Experiments have revealed that the devices have remarkable localization capabilities greatly exceeding resolution limits predicted from considerations of diffraction effects caused by the device aperture. Fluctuations in the imaged position of a distant point source appear to result from inhomogeneities in the non-ideal (water) medium. These fluctuation effects can be mitigated by data averaging until limits imposed by statistical non-stationarity of the medium are reached. In this talk, experiments measuring the fluctuation of point source images will be described and a theory of the resolution achievable by vector sensors will be presented.

10:05–10:20   Break

10:20

2aSA6. Spatial resolution of Helmholtz equation least-squares-based nearfield acoustical holography. Richard Dziklinski III and Sean Wu (Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

Traditional applications of near-field acoustical holography require that the acoustic pressures be taken at the Nyquist spatial sampling rate. For example, if the intended spatial resolution is 10 mm, then the spacing between measurement points should be 5 mm or less to ensure a target resolution. As a result, the total number of measurement points can be very high when a high spatial resolution in reconstruction is desired. The present paper examines the capability of using the Helmholtz equation least squares (HELS) method to reconstruct acoustic sources based on input data sampled at a rate lower than the Nyquist spatial sampling rate. Experimental results on two speakers separated by 9.53 mm are shown. A 4×4 microphone array is used with a spacing of 12.7 mm in the horizontal axis and 25.4 mm in the vertical axis, which is much larger than the spacing between the sound sources. Results show that if the coordinates of microphones are measured with a high accuracy (less than 2 mm) and the signal-to-noise ratio is 20 dB or higher, then these two sources can be reconstructed regardless whether the sounds are coherent or not. The reason for this phenomenon is explored in this paper.

10:35

2aSA7. Real time volumetric intensity visualization. Jacob Klos (Structural Acoust. Branch, NASA Langley Res. Ctr., Hampton VA 23681), Earl G. Williams (Naval Res. Lab., Washington DC 20375), Joel Tuss, and Bernard J. Sklanka (The Boeing Co., Seattle WA 98124)

Near-field acoustical holography (NAH) microphone array measurement technologies are currently being investigated by NASA, the Naval Research Lab (NRL) and Boeing. One type of array geometry under investigation is a spherical arrangement of microphones that enables the computation of the intensity field surrounding the sphere using spherical NAH theory. Due to the simplicity of the required computation and the ability to pre-compute several of the required quantities, the intensity field in a volume surrounding the sphere can be reconstructed and rendered in real time as a measurement is being made. Thus, sound sources can be identified in real time by roving the spherical array around in a sound field near suspected sources and tracking the sound intensity back to the source. This presentation demonstrates the capabilities of a real time spherical NAH system that has been developed by NASA, the NRL and Boeing. The topology of the software will be discussed and the capabilities of such a system will be demonstrated.

10:50

2aSA8. Numerical study of finite-amplitude source reconstruction. Micah R. Shepherd, Kent L. Gee (Dept. of Phys. and Astron. Brigham Young Univ. N-283 ESC Provo, UT 84606, micah.shepherd@byu.edu), and Mark S. Wochner, S. (Univ. of Texas, Austin, TX 78758)

Since its inception in the 1980s, near-field acoustic holography (NAH) has been widely used to determine structural source properties. More recently, it has also been used to determine aeroacoustic source behavior. Although several alternate forms of NAH have been developed, all are based on the (linear) Helmholtz equation and would presumably not function correctly for a finite-amplitude pressure wave propagating nonlinearly. However, when nonlinear effects or reconstruction distances are small, NAH may still be a useful tool for determining radiation characteristics of finite-amplitude sources such as jets and rockets. A two-dimensional numerical propagation scheme has been used to simulate simple noise propagation from an extended source to study the effect of nonlinearities on source reconstruction using NAH. Results from the study and their implication regarding the use of NAH for high-amplitude acoustic sources will be discussed.

11:05

2aSA9. Phased-array advancement for aeroacoustic measurements— A deconvolution approach for the mapping of acoustic sources (DAMAS). Thomas F. Brooks and William M. Humphreys, Jr. (NASA Langley Res. Ctr., Aeroacoustics Branch, M.S. 461, Hampton, VA 23681-0001, t.f.brooks@larc.nasa.gov)

Since the 1990s, there has been a significant increase in the use of phased arrays of microphones in studies of noise sources for both wind tunnel models and full-scale aircraft. This is in spite of the fact that array data interpretation has been burdened with considerable uncertainty when using traditional array processing. Results represent noise sources that are convolved with array beamform response functions, which themselves depend on array geometry, size, and frequency. Recently, Langley Research Center developed breakthrough methodology that decouples the array design and processing influence from the noise being measured, using a relatively simple and robust algorithm. The deconvolution approach for the mapping of acoustic sources (DAMAS) method removes beamforming characteristics from output presentations. This presentation shows results from several airframe noise studies and a MIT air brake noise study conducted at Langley's Quiet Flow Facility (QFF). It is shown that DAMAS can permit an unambiguous and accurate determination of acoustic source positions and strengths that cannot be otherwise attained. The acoustic community is accepting the methodology and applying it to other applications. At Langley, recent enhancements have been made to the original DAMAS code that allow the determination and separation of coherent and incoherent noise source distributions.

Session 2aSC

Speech Communication: Second Language Perception and Production (Poster Session)

Allard Jongman, Cochair
Univ. of Kansas, Dept. of Linguistics, Lawrence, KS 66045-2140

Rachel Hayes-Harb, Cochair
Univ. of Utah, Dept. of Linguistics, 255 S. Central Campus Dr., Salt Lake City, UT 84112-0492

Contributed Papers

All posters will be on display from 8:30 a.m. to 11:30 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 11:30 a.m.

2aSC1. Native English speakers and Arabic pharyngealization contrasts: Perceptual and acoustic data. Aleksandra M. Zaba (Univ. of Utah, 255 South Central Campus Dr., Rm. 2300, Salt Lake City, UT 84112)

Arabic contrasts nonpharyngealized (e.g., /t/) and pharyngealized (e.g., /T/, with a secondary pharyngeal articulation) consonants. These contrasts are manifested acoustically by native Arabic speakers primarily via manipulation of F2 onset and steady-state values for adjacent vowels (vowels adjacent to pharyngealized consonants have lower F2 values than those adjacent to nonpharyngealized consonants). Additionally, consonant duration, following vowel onset and steady-state F1 and F3 values, and burst duration have also been reported in the literature as relevant cues. Because vowels adjacent to pharyngealized consonants tend to have lower F2 values, native English listeners may be able to exploit their sensitivity to English front-back vowel contrasts to detect Arabic pharyngealization contrasts. For example, the vowel in /ta:/ may be perceived by English listeners as closer to English /ae/ and the vowel in /Ta:/ may be perceived as closer to English /a/. We report the results of discrimination and vowel identification experiments in native English listeners' ability to discriminate pharyngealization contrasts. These experiments examine the contributions of several acoustic cues, in particular those relevant to English vowel contrasts. We also report intertalker/dialect differences among Arabic speakers in the manipulation of acoustic cues to pharyngealization contrasts.

2aSC2. Perception of Japanese mora nasal /N/ and mora obstruent /Q/ by native Japanese and English speakers. Takeshi Nozawa (Program in Lang. Education, College of Economics, Ritsumeikan Univ., 1-1-1 Nojihigashi, Kusatsu Shiga 525-8577, Japan) and Elaina M. Frieda (Auburn Univ., 226 Thach Hall, Auburn, AL 36849)

Japanese mora nasal /N/ and more obstruent /Q/ have no definite point of articulation. For instance, /N/ becomes [m] when a bilabial consonant follows and [n] when an alveolar consonant follows. /Q/ is also realized as [p], [t], or [k], depending on the point of articulation of the following consonant. Four native speakers of Japanese produced /N/ and /Q/ in /CVNCV/ and /CVQCV/ contexts, where the consonant after /N/ and /Q/ is always a stop. Their utterances were recorded and digitized. The word final /CV/ was edited out, and the stimuli with the structure of /CVN/ and /CVQ/ were created. Twelve native speakers of Japanese and American English were recruited as listeners in Shiga, Japan and Auburn, AL, respectively. The American listeners were told to identify the word final consonant in a multiple-choice format. The Japanese listeners were told

that they were to tell whether /Q/ and /N/ were realized as [m], [n], [ŋ], and [p], [t], [k], respectively. The American listeners outperformed the Japanese listeners despite the fact that they had had no prior exposure to Japanese. This is probably because a stop does not occur or no phonemic contrast is allowed in a postvocalic position in Japanese.

2aSC3. Acquisition of lexical stress by Mandarin learners. Yuwen Lai and Joan Sereno (Dept. of Linguist., The Univ. of Kansas, Blake Hall, Lawrence, KS 66044)

Acoustic characteristics of stress were examined in second language learners' productions of English lexical stress. Fourteen minimal pairs of disyllabic English nouns and verbs contrasting in stress pattern (trochees and iambs; e.g., OBject and obJECT) were recorded. Eighteen Chinese learners of English (NNSs) (nine advanced, nine basic) and ten native speakers of English (NSs) participated in the experiment. Preliminary results indicate that subjects in all groups use vowel duration, F0, and intensity to distinguish the nouns and verbs but the correlates were realized differently. In nouns, NSs and both groups of NNSs pattern similarly in their use of duration. However, both basic and advanced non-native learners use F0 and intensity to a lesser extent than NSs. For verbs, NSs differ from NNSs in the realization of duration and F0 but not intensity. In addition, basic learners have significantly smaller duration differences but greater F0 differences between stressed and unstressed syllables when compared to NSs while advanced learners pattern in between the basic and NS groups. These results are discussed with respect to how suprasegmental features of the native language affect the prosodic acquisition of a second language.

2aSC4. The production of English corner vowels by English as a foreign language children learners in Taiwan. E-chin Wu (Grad. Inst. of Linguist., Natl. Taiwan Univ., No. 1, Sec. 4, Roosevelt Rd., Taipei, 10617 Taiwan)

This study aims to test whether EFL children learners in Taiwan would differentiate between the similar but not totally identical corner vowels /i/, /u/, and /a/ in English and Chinese. A picture naming task incorporated with a word guessing game was conducted on 12 subjects to collect vowel samples for acoustic analysis. The acoustic features being examined were the first and second formants. Results showed a tendency for the EFL children learners to distinguish between the Chinese and English corner vowels. This suggested that early EFL learners may possibly have separated sound systems created for different languages.

2a TUE. AM

2aSC5. English vowel training with different first-language vowel systems. Paul Iverson and Bronwen G. Evans (Dept. of Phonet. and Linguist., University College London, 4 Stephenson Way, London NW1 2HE, UK)

This study examined whether native speakers of Spanish and German learn differently when given auditory training for English vowels. Spanish has fewer vowels (5) than does German (18). Spanish speakers thus need to acquire more vowel categories when learning English than do Germans, but the relative sparseness of the Spanish vowel space may actually facilitate learning (i.e., there is more room for new categories). Prior to training, the Spanish and German speakers were matched between groups in terms of their English vowel identification accuracy. Each subject completed a battery of pre/post-training tests (e.g., English vowel identification in quiet and noise; perceptual mapping of best exemplars) and were given five sessions of high-variability vowel identification training. The results demonstrated that the German subjects improved more in their English vowel identification accuracy (average 20 percentage-point improvement) than did the Spanish subjects (average ten percentage-point improvement). It appears that, rather than interfering with new learning, the relatively dense German vowel space facilitated the formation of new vowel categories, perhaps by making German speakers more sensitive overall to categorical distinctions.

2aSC6. Training segmental productions to improve second language intelligibility. Deborah F. Burleson (Commun. Disord. Technol., Inc., 501 N. Morton St., #215, Bloomington, IN 47404 & Indiana Univ., Bloomington, IN 47404) and Jonathan Dalby (Indiana Univ.-Purdue Univ., Ft Wayne, IN 46805)

Training procedures that purport to improve the intelligibility of the speech of second language learners are often not validated by research. This study examines the effectiveness of English segmental intelligibility training for native Mandarin Chinese-speaking learners. Ten Mandarin Chinese speakers trained on 16 American English phonemic contrasts. Half of the participants trained on 16 vowel errors common in Mandarin-accented English and the other half on 16 common consonant errors. Each participant received 15 h of computer-based, minimal word-pairs training. Native American English listeners judged the intelligibility of pre- and posttraining recordings of isolated words read by the trainees in a two-alternative, forced-choice identification task. Listeners also judged the intelligibility of sentences produced by trainees and recorded pre- and post-training. Results showed that training produced significant improvement in the intelligibility of isolated words included in the training and also in phonologically similar words not included in training. Vowel and consonant training participants made similar word-level improvement. The effect of word-level training on improvement in sentence intelligibility, while small, was also significant when results from participants in both vowel and consonant training groups were considered together. These results encourage continued exploration of structured computer-assisted pronunciation training to improve second language intelligibility.

2aSC7. Nonphonetic variability and speech categorization by native and non-native learners. Betty Tuller and McNeel Jantzen (Ctr. for Complex Systems and Brain Sci., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431, tuller@ccs.fau.edu)

Natural speech stimuli used in studies of phonological learning typically include several in talkers and phonetic environments because nonphonemic variability usually facilitates learning [e.g., Lively *et al.*, J. Acoust. Soc. Am. (1993)]. In contrast, our previous work demonstrated that perceived talker variability in a synthetic test set markedly decreased the ability of monolingual American English subjects to distinguish between the voiced dental stop consonant and the voiced alveolar stop [d] [Jantzen and Tuller, J. Acoust. Soc. Am. (2006)]. Only the latter is phonemic in American English. There are several possible explanations for this that are currently being evaluated. Here we examine the effect of perceived talker variability on Malayalam listeners' identification of three

dental to alveolar continua that differ in F0. In Malayalam, both the dental and the alveolar stop consonants are phonemic. The results shed light on whether talker information and phonetic information share a representational structure. [Work supported by NSF and ONR.]

2aSC8. The Wildcat corpus of native- and foreign-accented English. Ann R. Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60201), Rachel E. Baker, Arim Choi, Midam Kim, and Kristin J. Van Engen (Northwestern Univ., Evanston, IL 60201)

For a wide range of socio-political reasons, many conversations across the globe today are between interlocutors who do not share a mother tongue. As a resource for investigating the nature and broad implication of speech communication in a global context, we are currently developing a large database of speech recordings from native and non-native speakers of English. A key feature of this database is that, in addition to providing recordings of scripted materials, the speakers are recorded in pairs (all possible pairing of native and non-native English speakers) as they work together on a novel, interactive, goal-oriented task. The final corpus will include fully segmented and phonetically aligned digital recordings of both the scripted and unscripted speech samples along with complete orthographic transcriptions. We are currently using a first version of this database to track speaker-listener alignment over the course of a conversation, to compare phonetic features of speech addressed to native versus to non-native speakers, and to assess communication efficiency across various native and non-native speaker pairings (e.g., how long does it take for a team-based task to reach successful completion when one, both, or neither of the team members is a non-native speaker?)

2aSC9. Asymmetrical costs of code-switching in Mandarin-Min bilingual speakers. YiHsuan Huang and Janice Fon (Grad. Inst. of Linguist., Natl. Taiwan Univ., No. 1, Sec. 4, Roosevelt Rd., Taipei, 10617 Taiwan, r94142011@ntu.edu.tw)

This paper aims to explore the processing cost incurred by code-switching for bilingual speakers. Twenty Mandarin-Min bilinguals performed a picture-naming task, in which hand-drawn pictures were color-coded for the two languages, and switching points were predetermined but variable. Half of the subjects acquired Min prior to Mandarin and half of them learned both simultaneously as their L1. Forty target stimuli were divided into two groups. One was the cognate group, in which the target stimuli form cognates in the two languages (e.g., Mandarin: jing3-cha2 versus Min: keng3-chhat4 police). The other was the noncognate group, in which the target stimuli do not form cognates in the two languages (e.g., [Mandarin] sha1-fa1 versus [Min] phong3-i2 sofa). Preliminary results showed that the response latencies of naming in code-switching trials were larger than nonswitching ones. Moreover, asymmetry time costs were found in different switching directions between the two languages. Naming latencies were longer when subjects switched from Mandarin to Min than when they switched from Min to Mandarin. In addition, reaction time for code-switching of noncognate words tended to be longer than that for cognate ones. Furthermore, subjects who acquired Min first reacted more rapidly than those who acquired both languages simultaneously.

2aSC10. Individual differences in the production and perception of the voicing contrast by native and non-native speakers of English. Bruce L. Smith (Dept. of Commun. Sci. and Disord., 390 S. 1530 E., Rm. 1201 BEHS, Univ. of Utah, Salt Lake City, UT 84112, bruce.smith@hsc.utah.edu) and Rachel Hayes-Harb (Dept. of Linguist., Univ. of Utah, Salt Lake City, UT 84112)

Investigations of acoustic characteristics of speech production and their perceptual correlates often determine that various common patterns are observed among groups of talkers/listeners. It is also sometimes reported that a certain amount of noise occurred in the data, i.e., that not all subjects fit general patterns that were observed. While individuals who do

not demonstrate typical patterns are often viewed as outliers, subjects who represent exceptions also provide important information. The present study reports findings obtained from 15 native speakers of English and 15 non-native (native Mandarin) speakers of English who were recorded producing several final-position voiced/voiceless CVC word pairs and who also made voicing judgments about voiced/voiceless cognates produced by other native English and native Mandarin speakers. All native subjects in the present study produced multiple acoustic cues in making voicing distinctions (e.g., vowel duration, final stop closure duration, voicing during final stop occlusion, release burst); however, not all subjects used all these acoustic cues or the same combinations of cues in their voicing contrasts. Only about half the native Mandarin speakers used multiple acoustic cues in attempting to produce a voicing contrast. Various group trends, as well as subgroup variations, in production and perception will be discussed.

2aSC11. Production of lexical stress by Spanish speakers of English as a second language and its effect on perceptual judgments of intelligibility and accentedness. Paul Edmunds (Dept. of Linguist., Univ. of New Mexico, 526 Humanities, Albuquerque, NM 87131)

Acoustic analyses of recordings of read speech were used to compare the production of lexical stress in multisyllabic words by Spanish ESL (English as a Second Language) speakers to that of native English speakers. The participants read short sentences containing multi-syllabic target words for which syllable duration, intensity, and fundamental frequency in the pronunciations of stressed and unstressed syllables were compared. Target words were equally divided based on syllable count, the syllable that should receive lexical stress based on native English pronunciation, and Spanish/English cognate status. Both orthographically identical cognates (e.g., radio) and orthographically non-identical cognates (e.g., necessary/necesario) were included. Average values of syllable duration, intensity and fundamental frequency were recorded across the speaker groups, and effects of cognate status on production were noted. Target words pronounced by the native speakers (to avoid differing segmental factors) were then resynthesized using the PSOLA algorithm to simulate the acoustic patterns of the ESL speakers. Native English listeners were asked to rate a set of stimuli embedded in a carrier phrase for intelligibility and accentedness. Results of the listening tasks suggest that variations in the acoustic properties under investigation may impact a listeners perception of speech intelligibility and accentedness.

2aSC12. Role of orthographic information in the auditory processing of novel L2 words. Paola Escudero (Inst. of Phonetic Sci., Univ. of Amsterdam, Amsterdam, The Netherlands, escudero@uva.nl), Rachel Hayes-Harb (Univ. of Utah, Salt Lake City, UT 84112), and Holger Mitterer (Max Planck Inst. for Psycholinguist., Nijmegen, The Netherlands)

The English auditory word recognition of proficient Dutch-English bilinguals was examined in a follow-up study to Weber and Cutler (2004), who found, using eye-tracking technology, that these listeners could differentiate English /E/ and /ae/ in their lexical representations despite not being able to differentiate them in a perception task. The present study used the same eye-tracking method to test the hypothesis that knowledge of the spelled forms of words may allow learners to establish contrastive lexical representations for words containing perceptually difficult L2 contrasts. Fifty Dutch-English bilinguals learned nonword labels for nonobjects; some of the nonwords contained /E/ and some /ae/. One-half of the subjects were exposed to auditory forms and pictures only, while another set of subjects was additionally provided spelled forms of the words (the letters e and a differentiate /E/ and /ae/ words). At test, only subjects who received orthographic information during the word learning phase exhibited the same pattern found in Weber and Cutler (2004), providing support for the hypothesis that orthography can contribute to contrastive lexical representations, even for contrasts that are perceptually neutralized. The implications of these results for theoretical and applied issues in L2 speech processing will be presented.

2aSC13. Context effects in recognition of German disyllabic words and nonwords by native and non-native listeners. Robert A. Felty (Depts. of Linguist. and German, Univ. of Michigan, 3110 MLB, 812 E. Washington St., Ann Arbor, MI 48109-1275, robfelty@umich.edu)

Thirty-two native (L1 = German) and 30 non-native (L1 = English) listeners heard German CVCCVC words and nonwords mixed with noise. Of the words, half were monomorphemic and half bimorphemic. The j-factor model [Boothroyd and Nittrouer, J. Acoust. Soc. Am. **84**, 101–114 (1988)] was used as a measure of lexical context effects. For both native and non-native speakers, words showed greater context effects than nonwords, though the difference was not as large for non-native speakers. Monomorphemic words also exhibited greater context effects than bimorphemic words, again with a larger effect for native speakers. For both native and non-native listeners, neighborhood density had a significant effect, although the amount of variation explained was much less for non-native listeners. This is interpreted to be a result of the smaller lexicon of the non-native listeners. Misperceptions between native and non-native speakers were also compared, with non-native speakers showing patterns predictable by phonological differences between the two languages. The j-factor results extend previous research using CVC stimuli, providing additional support for its efficacy as a measure of context effects. The j-factor results also provide a new finding, in that j_{word} does not seem to scale linearly with word length. [Work supported by NIH/NIDCD.]

2aSC14. Speeded discrimination of American vowels by experienced Japanese late L2 learners. Kikuyo Ito, Franzo F. Law II, Mieko N. Sperbeck, Shari Berkowitz (Speech Acoust. and Percept. Lab., City Univ. of New York Grad. Ctr., 365 Fifth Ave., New York, NY 10016-4309), Yana D. Gilichinskaya, Marisa Monteleone, and Winifred Strange (City Univ. of New York Grad. Ctr., New York, NY 10016-4309)

This study is part of a larger project to examine online perception of American (AE) vowels by non-native listeners from languages with small vowel inventories. A speeded ABX categorial discrimination task with disyllables /Vpə/ was used to examine Japanese L2 learners relative difficulties differentiating 8 AE vowels contrasted in 13 pairs: 6 adjacent height contrasts among front [i:/ɪ, ɪ/[g\], [g\]/æ:] and back [u:/ʊ, ʊ/ʌ, ʌ/ɑ:] vowels and 3 front/back pairs [ɪ/ʊ, [g\]/ʌ, æ:/ɑ:]; 4 nonadjacent height pairs [i:/[g\], ɪ/æ:, u:/ʌ, ʊ/ɑ:] served as controls. Reaction times for native AE listeners did not vary across the 13 contrasts, despite differences in psychoacoustic distance between vowel pairs, reflecting automatic processing of native contrasts. However, for Japanese listeners, performance on contrasts among psychoacoustically more similar [[g\], æ:, ɑ:, ʌ] yielded the slowest latencies, especially for front/back pairs that did not differ in duration. Height contrasts between high/mid-high vowels (height + duration) and mid-high/mid-low vowels (height only) were discriminated somewhat faster, although not as fast as control contrasts. The latter were discriminated as rapidly as native Japanese height and length contrasts. Results suggest that response latencies are a sensitive measure of differences in automaticity of online processing of native and non-native contrasts.

2aSC15. The effect of phonetic similarity on L2 speakers in discerning L1 and L2 consonants. Sally Chen and Janice Fon (Grad. Inst. of Linguist., Natl. Taiwan Univ., 1, Roosevelt Rd., Sec. 4, Taipei, 106 Taiwan)

This study investigated how L1 and L2 sounds of different degrees of similarity are perceived by L2 speakers. Twenty-six participants were tested with six Mandarin-English C-/i/ syllable pairs varying in degrees of similarity in the initial consonant. Two competing models were examined, the speech learning model (Flege, 1995) and the perceptual assimilation model (Best, 1995). The former proposes that the greater the perceived dissimilarity between an L2 sound and its closest L1 counterpart, the easier it is for L2 speakers to discern the difference, while the latter proposes that L2-to-L1 assimilation corresponds to their articulatory distance. As for two L2 sounds corresponding to the same L1 target, perceptual

performance may still differ depending on how similar the L2 sounds are to the target. Results showed that perceptual performance was independent of L2 proficiency. The SLM is supported in that the accuracy for discerning individual consonants of "dissimilar" pairs was higher than that for "similar" pairs, which in turn was higher than that for "identical" pairs. However, the accuracy for discerning "identical" pairs was still significantly higher than the chance level, indicating that L2 speakers are still able to tell the difference, which can be better explained by the PAM.

2aSC16. On the perception of Korean and English sibilants in second language acquisition. Sang Yee Cheon (Dept. of East Asian Lang. and Lit., Univ. of Hawaii at Manoa, 1890 East West Rd., 382 Honolulu, HI 96822, scheon@hawaii.edu)

This study investigates the effect of phonetic similarity between L1 and L2 sounds on the perception of L2 contrasts. It employs hypotheses from Flege's speech learning model as well as two types of assimilation patterns from Best's perceptual assimilation model. L1 L2 phonetic similarities were measured perceptually and acoustically prior to the experiments. The first experiment examined the perception of the Korean voiceless sibilant contrast /s-s*/ by American English L2 learners. The second experiment examined the perception of the English voiceless sibilant contrast /s-ʃ/ by Korean L2 learners. The findings of the first experiment show that a higher degree of phonetic similarity did not always result in good identification of L2 contrasts by American English advanced learners of Korean; overall, however, more similar L2 sounds were perceived better than less similar L2 sounds. In the second experiment, L2 contrasts that are acoustically and perceptually different from each other were perceived by L2 learners of Korean just as well as native speakers of American

English perceived them. The results were not compatible with the predictions of either Best's or Flege's models, but the second result partially supported Best.

2aSC17. Cross-language perception of word-final stops: A comparison of Cantonese, Japanese, Korean, and Vietnamese listeners. Kimiko Tsukada (Dept. of East Asian Lang. and Lit., Univ. of Oregon, Eugene, OR 97403), Shunichi Ishihara (The Australian Natl. Univ., Australia), Thu T. A. Nguyen (Univ. of Queensland, Australia), and Rungpat Roengpitya (Mahidol Univ., Thailand)

Four groups of Asian language speakers [Cantonese (NC), Japanese (NJ), Korean (NK), Vietnamese (NV)] were compared for their accuracy to discriminate English and Thai word-final stop contrasts (/p-t/, /p-k/, /t-k/). Word-final stops can be either audibly released or unreleased in English, whereas they are always unreleased in Thai. English (but not Thai) was familiar to all participants. In Cantonese, Korean, and Vietnamese, word-final stops are unreleased as in Thai. Japanese, on the other hand, does not permit word-final stops. All groups were more accurate in discriminating English than Thai contrasts. However, their response patterns differed according to their first languages (L1). NK was almost as accurate in Thai as in English while NJ was much better in English than in Thai. NC and NV were intermediate between NK and NJ. These findings suggest that experience with specific (i.e., unreleased) and native phonetic realization of sounds may be essential for accurate discrimination of word-final stop contrasts. However, in order to explain the differences between NC, NK, and NV in their discrimination of unreleased Thai stops, it is necessary to examine the robustness of phonetic cues to word-final stops in their L1s. The role of L1 experience in cross-language perception is discussed.

TUESDAY MORNING, 5 JUNE 2007

ALPINE EAST, 8:25 A.M. TO 12:00 NOON

Session 2aUW

Underwater Acoustics: Propagation and Modeling

Jon Collis, Chair
Boston Univ., Boston, MA 02215

Chair's Introduction—8:25

Contributed Papers

8:30

2aUW1. Propagation in an elastic wedge using the virtual source technique. Ahmad T. Abawi and Michael B. Porter (Heat, Light, and Sound Res., Inc., San Diego, CA 92130, Ahmad.Abawi@HLSRes.com)

In this paper the virtual source technique is used to model propagation of waves in a range-dependent ocean overlying an elastic bottom with arbitrarily-shaped ocean-bottom interface. The virtual source technique provides the means to impose boundary conditions on arbitrarily-shaped boundaries by replacing them by a collection of sources whose amplitudes are determined from the boundary conditions. The method is applied to propagation in an elastic Pekeris waveguide, an acoustic wedge and an elastic wedge. In the case of propagation in an elastic Pekeris waveguide, the results agree very well with those obtained from the wavenumber integral technique, as they do with the solution of the parabolic equation (PE) technique in the case of propagation in an acoustic wedge. The results for propagation in an elastic wedge qualitatively agree with those obtained from an elastic PE solution.

8:45

2aUW2. Scattering of a Bessel beam by a sphere: Geometric interpretation and resonance excitation. Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164-2814)

The exact solution for the scattering of sound by an isotropic sphere centered on a Bessel beam in an inviscid fluid was recently given [P. L. Marston, J. Acoust. Soc. Am. **121**, 753–758 (2007)]. The solution gives insight into the scattering by a broader class of beams and provides a benchmark for the testing of finite-element based scattering algorithms. An important parameter in the Bessel beam solution is the cone angle of the Bessel beam which is related to the ratio of the beam size to the sphere radius and to the wave-number-radius product. For an impenetrable sphere there is a simple geometric interpretation of the scattering pattern verified by quantitative ray theory. The modification of the partial wave series due to Bessel beam illumination may be interpreted using the Van de Hulst localization principle. In the scattering by elastic spheres, specific partial-waves may be suppressed by appropriate selection of the cone angle of the beam. Compared with the plane-wave case, this suppression of partial

waves may increase or decrease the backscattering, depending on the situation. [Work supported by the Office of Naval Research.]

2aUW3. Regions that influence acoustic propagation in the sea at moderate frequencies, and the consequent departures from the ray-acoustic description. John Spiesberger (Dept. Earth and Env. Sci., Univ. of Pennsylvania, Philadelphia, PA 19104-6316, johnsr@sas.upenn.edu)

In the limit where a transient signal is comprised of very large frequencies, spatial regions within an inhomogeneous medium that influence the propagation from a source to a receiver lie along one or more ray paths. At lower frequencies for which the geometrical acoustic approximation is of borderline applicability, the regions that influence such transient signals are extended because of diffraction. Previous research has addressed the numerical determination of those spatial regions that influence propagation at low frequency. The present paper addresses the question of how high the center frequency need be so that the regions of influence are nearly described as ray paths for a model ocean in which the speed of sound increases nearly linearly with depth from a perfectly reflecting surface. Computations indicate that near 2500 Hz and at a range of 50 km, the region of influence resembles a ray. Noticeable departures from the ray picture are found at a range of 500 km. Various physical and mathematical causes for the departures from the ray propagation model for lower frequencies and for greater ranges are identified and discussed. [Work supported by the ONR.]

2aUW4. Three-dimensional coupled-mode model of propagation and scattering around a conical oceanic seamount. Wenyu Luo and Henrik Schmidt (Massachusetts Inst. of Technol., Cambridge, MA 02139)

Motivated by interesting anomalies observed in recent transmission experiments across the Kermit Seamount in the Pacific, a numerically efficient three-dimensional (3D) propagation and scattering model has been developed based on 3D coupled mode theory for axisymmetric bathymetry. The 3D coupled mode approach applied here is based on the same spectral representation of the field as earlier models presented by Athanassoulis and Propathopoulous [J. Acoust. Soc. Am. **100** (1996)] and Tarudakis [J. Comp. Acoust. **4** (1996)]. However, the earlier formulations were severely limited in terms of frequency, size, and geometry of the seamount, the seabed composition, and the distance of the source from the seamount, and are totally inadequate for modeling the Kermit experiment. By introducing a number of changes in the numerical formulation and using a standard normal mode model (CSNAP) for determining the fundamental modal solutions and coupling coefficients, orders of magnitude improvement in efficiency and fidelity have been achieved, allowing for realistic propagation and scattering scenarios to be modeled, including effects of seamount roughness, realistic sedimentary structure. By comparing to traditional n-by-2D coupled mode results, the model is used to demonstrate and investigate the surprisingly strong significance of 3D effects associated with propagation across oceanic seamounts such as Kermit, reaching up to the depth of the SOFAR channel. [Work supported by the US Office of Naval Research, Code 321OA.]

2aUW5. Low-frequency resonances from the circumnavigation of waves on plastic cylinders: Ordinary and evanescent-wave excitation. Aubrey L. Espana and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

Resonances of cylinders provide useful information for separating signatures of targets of interest from the clutter of seafloor echoes. In prior work, the coupling of evanescent waves to low-frequency axial resonances of cylinders was investigated for targets in a laboratory-based simulated fluid sediment [C. F. Osterhoudt and P. L. Marston, J. Acoust. Soc. Am. **120**, 3143 (2006)]. To facilitate investigations of the coupling to resonances associated with guided-wave circumnavigation, it is desirable to

identify small cylindrical targets having appropriate regular resonance features visible with plane-wave illumination. The present investigation concerns such resonances on small polystyrene cylinders in water. Analogous features were previously modeled for acrylic spheres [B. T. Hefner and P. L. Marston, J. Acoust. Soc. Am. **107**, 1930–1936 (2000)]. Those features were associated with the generalization of a Rayleigh wave on the solid. (Some other authors prefer to label these features as Scholte-Stoneley waves.) Measurements of the free-field acoustic backscattering by a small truncated polystyrene cylinder reveal the presence of analogous resonances. The measurements facilitate a comparison with approximations discussed by Stanton [J. Acoust. Soc. Am. **83**, 64–67 (1988)]. Resonance excitation by evanescent waves was also confirmed. [Work supported by the Office of Naval Research.]

2aUW6. Benchmarking the variable rotated and mapping parabolic equation solutions. Jon M. Collis (Boston Univ., Boston, MA 02215) and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

Improving the capability of handling variable sloping interfaces and boundaries with the parabolic equation method is an active area of research. The mapping solution [M. D. Collins, *et al.*, JASA **107**], and the variable rotated parabolic equation [D. A. Outing *et al.*, JASA **115**] are recent advances in this area. Both solutions handle variable bathymetry by approximating in terms of a series of regions of constant slope. The mapping solution transforms the bottom to a line, and the parabolic equation is solved in a coordinate system aligned with the interface. At junctions between regions, a phase correction is applied to the field. The variable rotated solution rotates the coordinate system in each region so that the parabolic equation is solved in a system that is aligned with the boundary. The junctions between regions are treated by overshooting the point where the slope changes and interpolating and extrapolating to obtain the field in the next region. Although both methods handle variable bathymetry, the variable rotated solution should provide greater accuracy, especially when there are relatively large slopes. The two solution methods are benchmarked against each other and an analytic solution. [Work supported by the Office of Naval Research.]

2aUW7. Wavefront modelling of pulse propagation. Chris Tindle (Phys. Dept., Univ. Auckland, Auckland, New Zealand), Grant Deane (Scripps Inst. Oceanogr., La Jolla, CA 92093), and James Preisig (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Wavefront modelling has been developed for the rapid calculation of receiver waveforms in underwater sound pulse propagation. The method uses a solution of the wave equation which expresses the field as a sum over phase integrals. Each phase integral corresponds to a particular sequence of reflections or turning points and is evaluated analytically by using stationary phase approximations. Each point of stationary phase corresponds to a ray path. Pairs or triplets of stationary phase points correspond to caustics or foci and the phase and amplitude of the field is found directly. The method can handle rapid range dependence such as occurs with surface waves in shallow water and double scattering from the same surface. The method is fast and accurate and examples will be discussed. [Work supported by ONR.]

2aUW8. Probability density functions of acoustic modal amplitudes by polynomial chaos expansions. Kevin D. LePage (Naval Res. Lab., Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375)

Polynomial chaos (PC) expansions have recently been derived to model the statistical behavior of complex modal amplitudes in uncertain waveguides. Here we explore the statistical trajectories of complex modal amplitudes in phase space for one and two degrees freedom (DOF) of waveguide uncertainty. It is shown that for one DOF of uncertainty, the

modal amplitudes fall along a trajectory in phase space, while for higher DOFs the phase space of the amplitudes is space filling. Results are compared to phase space characteristics of the adiabatic and saturated limits of waveguide uncertainty, and a discussion of the range progression of the scintillation index of modal amplitude is undertaken. Comparisons between PC estimates of the probability density function of the modal amplitudes and Monte Carlo histograms are also made. [Work supported by ONR.]

10:45

2aUW9. Frequency variability of modal attenuation coefficients. Wendy Saintval, William L. Siegmann (Rensselaer Polytechnic Inst., Dept. of Mathematical Sci., 110 8th St., Troy, NY 12180), William M. Carey, and Allan D. Pierce (Boston Univ., Boston, MA 02115)

The intrinsic sediment attenuation in adiabatic waveguides affects transmission loss through the modal attenuation coefficients (MACs). For sand-silt bottoms, the non-linear frequency dependence of attenuation in the upper sediment layer significantly affects the frequency variation of the MACs. This variation is shown to depend strongly on the water, and weakly on the upper sediment layer, sound-speed profiles. The MACs are calculated for a layered environmental model that permits good comparisons with values derived from experimental data in the Gulf of Mexico [Ferris, JASA, **82** (1972)]. The comparisons illustrate the dependence on measured water profiles. The rate of increase of transmission loss with range is conveniently measured by an effective attenuation coefficient. The relationship between this quantity and the MACs is known when one propagating mode dominates [Evans and Carey, JOE, **23** (1998)]. We describe the interesting connection for downward refracting profiles at mid-frequencies with multiple propagating modes. [Work partially supported by ONR.]

11:00

2aUW10. A robust method for approximating acoustic field uncertainty in underwater sound channels. Kevin R. James and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, 2212 G. G. Brown, 2300 Hayward Ave., Ann Arbor, MI 48109)

Environmental uncertainty in underwater sound channels is often the dominant source of error in acoustic field predictions. Standard Monte Carlo techniques to quantify the effect of this uncertainty are prohibitively time consuming for many applications. This presentation evaluates an alternative method for approximating the relationship between uncertain environmental parameters and uncertainty in the predicted field. Using FFT techniques, the effect of small changes in uncertain parameters is correlated to small spatial displacements of the original field. By appropriately extrapolating this relationship between uncertain parameters and the resulting field, an approximate probability distribution can be calculated for the field amplitude. A variety of test cases are shown to evaluate the accuracy of this method in underwater sound channels. Several uncertain parameters are addressed, including bottom characteristics, water column depth, source characteristics, and water column properties. Application of this method to current field prediction routines is discussed. Results are compared to exact solutions, Monte Carlo simulations, and simple linear approximations. In particular, for the sound channels studied, when environmental uncertainties remain within meaningful limits, approximate field amplitude distributions generally exhibit means and standard deviations within 10% of those calculated from the exact field amplitude distribution. [Work sponsored by the Office of Naval Research, Code 321OA.]

11:15

2aUW11. Acoustic propagation in uncertain waveguides with spatially varying wave speeds. John Burkhardt (US Naval Acad. 590 Holloway Rd. Annapolis MD 21402, burkhard@usna.edu)

Acoustic propagation in the littoral is a complex process involving acoustic interaction with the ocean surface, the ocean bottom, volume scattering as well as variable propagation characteristics throughout the volume. When modeled stochastically this problem has been approached using a variety techniques including the Monte Carlo method, Neumann expansions and the stochastic finite element method. The last of these has found considerable use because it offers computational efficiencies over the Monte Carlo method and fewer modeling restrictions than the Neumann expansions. This presentation will discuss the development of a one-dimensional stochastic finite element formulation that accounts for spatial variability in wave speed, an area of intense interest not currently addresses in the literature.

11:30

2aUW12. Comparison of two and three spatial dimensional solutions of the wave equation at ocean-basin scales in the presence of internal waves. John Spiesberger (Dept. Earth and Env. Sci., Univ. of Pennsylvania, Philadelphia, PA 19104-6316, johnsr@sas.upenn.edu)

Numerical solutions are given for a parabolic approximation to the acoustic wave equation at 75 Hz in two and three spatial dimensions to determine if azimuthal coupling of the field significantly affects horizontal coherence. Coupling is a small effect at 4000 km in the presence of internal gravity waves. This implies that accurate solutions are possible using computations from uncoupled vertical slices through the field. The shape of horizontal coherence is inconsistent with shapes given by two theories. Estimates of horizontal coherence at 4000 km and 25, 50, and 75 Hz are 10, 2, and 1 km respectively. [Work supported by the ONR.]

11:45

2aUW13. Modeling marine mammal sound exposure maps for seismic exploration airgun arrays. Natalia A. Sidorovskaia (Phys. Dept., UL Lafayette, UL BOX 44210, Lafayette, LA 70504-4210, e-mail: nas@louisiana.edu), Arslan M. Tashmukhambetov, George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Ocean seismic surveys rely on acoustic pulses produced by seismic airgun arrays for mapping ocean bottom layers in search of hydrocarbon indicators. Seismic exploration arrays produce mainly low frequency acoustic energy with a small amount of higher frequency content. It is known that the energy distribution in the ocean depends on both the array and the propagation-channel characteristics. A main concern for marine mammals is possible energy refocusing in depth and range due to waveguide propagation. Standard ocean propagation models are combined with industry airgun modeling packages to simulate the three-dimensional broadband acoustic field in the frequency and time domains for a seismic exploration array in typical Gulf of Mexico environments, including seasonally formed surface ducts. To emulate the accepted acoustic energy integration time for sperm whales, three-dimensional sound exposure maps are generated using a 200-msec integration window for transient seismic pulses. The maps are verified by comparison to available experimental data collected by the Littoral Acoustic Demonstration Center in a 2003 seismic calibration experiment. The proposed workflow can be integrated into procedures for predicting the environmental effects of seismic exploration surveys in waveguides. [Research partially supported by the Industry Research Funding Coalition through the International Association of Geophysical Contractors.]

Meeting of the 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,
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
2117 Robert Drive, Champaign, IL 61821

V. Nedzelnitsky, U.S. Technical Advisor (TA) for IEC/TC 29, Electroacoustics
National Institute of Standards and Technology (NIST), Sound Building, Room A147, 100 Bureau Drive, Stop 8221, Gaithersburg, MD 20899-8221

Three of the four Accredited Standards Committees and three of the nine U.S. Technical Advisory Groups administered by the Acoustical Society of America will meet in conjunction with the 153rd meeting of the Acoustical Society of America at **The Hilton Salt Lake City Center, Salt Lake City, Utah** from **4−8 June 2007**. The specific meeting details are shown below. Additional details regarding lodging, transportation, etc. can be found on the Acoustical Society of America's web site at http://asa.aip.org. (For details about the meeting of Accredited Standards Committee S2, Mechanical Vibration and Shock and the U.S. TAGs to ISO/TC 108 and its subcommittees, please visit the standards page on the web site.)

The reports of the Chairs of these TAGs will not be presented at any other S Committee meeting.

The meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards Committees S1, S3 and S12, which are scheduled to take place in the following sequence:

ASC S12 Noise	**5 June 2007**	**2:00 p.m. to 5:00 p.m.**
ASC S1 Acoustics	**6 June 2007**	**8:30 a.m. to 10:15 a.m.**
ASC S3 Bioacoustics	**6 June 2007**	**10:30 a.m. to 12:00 p.m.**

Discussion at the Standards Committee Plenary Group meeting will consist of national items relevant to all S Committees.

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 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	S1 and S3
P. D. Schomer, Chair	**ISO/TC 43/SC1** Noise	S12
IEC		
V. Nedzelnitsky, U.S. TA	**IEC/TC 29** Electroacoustics	S1 and S3

Informational note: The meetings of Accredited Standards Committee S2, Mechanical Vibration and Shock and the U.S. TAGs to ISO/TC 108 and its subcommittees will be held on Tuesday, 19 June 2007 at 1:00 p.m. in San Antonio, TX, in conjunction with the meeting of the Vibration Institute. For details, contact the Secretariat or visit the web site http://asa.aip.org and click on ''Standards Info.''

Session 2pAAa

Architectural Acoustics and the National Council of Acoustical Consultants: Student Design Competition

Robert D. Coffeen, Cochair
Univ. of Kansas, 134 Marvin Studios, Lawrence, KS 66045

Byron W. Harrison, Cochair
The Talaske Group Inc., 105 N. Oak Park Ave., Oak Park, IL 60301

Norman Phillip, Cochair
Yantis Acoustical Design, 720 Olive Way, Seattle, WA 98101

The Technical Committee on Architectural Acoustics of the Acoustical Society of America, the National Council of Acoustical Consultants, and the Robert Bradford Newman Student Award Fund are sponsoring the 2007 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 a music performance hall and related facilities that will be used primarily for a college opera program. 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 2pAAb

Architectural Acoustics and Noise: The Technical Committee on Architectural Acoustics Vern O. Knudsen Distinguished Lecture

Lily M. Wang, Chair
Univ. of Nebraska, Lincoln, Architectural Engineering, 101A Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681

Chair's Introduction—4:00

Invited Paper

4:05

2pAAb1. The soundscape of modernity: Architectural acoustics and the culture of listening in America, 1900−1933. Emily Thompson (Dept. of History, Princeton Univ., 136 Dickinson Hall, Princeton NJ 08544-1174, emilyt@princeton.edu)

The American soundscape changed dramatically during the early decades of the twentieth century as new acoustical developments transformed both what people heard and the ways that they listened. What they heard was a new kind of sound that was the product of modern technology. They listened as newly critical consumers of aural commodities. Reverberation equations, sound meters, microphones, and acoustical tiles were deployed in places as varied as Boston's Symphony Hall, New York's office skyscrapers, and the sound stages of Hollywood. The result was that the many different spaces that constituted modern America began to sound alike—clear, direct, efficient, and non-reverberant. While this new modern sound said little about the physical spaces in which it was produced, it has much to tell us about the culture that created it. This talk will explore the history of modern sound and modern culture in early twentieth-century America.

Session 2pAB

Animal Bioacoustics: Seismic Communication in Animals

Peggy Hill, Cochair
Univ. of Tulsa, 600 S. College, Tulsa, OK 74104

Caitlin O'Connell-Rodwell, Cochair
Stanford Univ., Dept. of Otolaryngology, Head & Neck Surgery, 801 Welch Rd., Stanford, CA 94305

Invited Papers

1:00

2pAB1. A snapshot of known animal communication via the vibration channel. Peggy Hill (Faculty of Biological Sci., The Univ. of Tulsa, 600 South College, Tulsa, OK 74104, peggy-hill@utulsa.edu)

When a new field, or sub-discipline, first emerges in the scientific arena, terminology is often borrowed from mostly compatible and closely related fields, or those fields from which the new one is emerging. Jargon is employed within the new field when the established vocabulary fails. A point is reached when the best interest of communication with those outside the new field requires development of a new vocabulary, or at least a specific redefinition of broadly used terms, so that the new field can continue to develop and fill its own niche. Animals from fiddler crabs to elephants communicate via the vibration channel. In fact, when researchers have suspected substrate borne signaling, it has almost always been confirmed. In this presentation, I will provide examples from the literature of animals across taxa that are known to communicate in this way, and the contexts within which signals are sent and received. Further, I will pose questions (with few answers) concerning selection for use of the channel, along with examples of instances where changes in the use of terminology over a couple of decades has restricted access to early work in this field.

1:30

2pAB2. Host shifts and the evolution of vibrational communication in treehoppers. Reginald B. Cocroft (Biological Sci., Univ. of Missouri, Columbia, MO 65211, CocroftR@missouri.edu)

Speciation in plant-feeding insects is associated with shifts to novel host plants. Along with differences in a suite of life history and fitness-related traits, closely related insect species that live on different host plants often differ dramatically in their communication systems. Differences in traits involved in sexual communication can provide a source of reproductive isolation and thereby facilitate speciation. Here the relationship between host shifts and changes in sexual communication is investigated in a group of host-specialist insects, the *Enchenopa binotata* species complex of treehoppers (Hemiptera: Membracidae). Each of the eleven species in the complex has unique plant-borne signals used in the process of pair formation. Variation between species in female preferences and male signal traits indicates that closely-related species on different host plants have experienced divergent sexual selection. Changes in host use can also promote signal evolution through divergent natural selection on signal form and through phenotypic plasticity. Host shifts can thus have multiple effects on the evolution of communication systems in plant-feeding insects. [Work support for this research was provided by NSF.]

2:00

2pAB3. Changes in host plant use favor divergence of vibrational signals in treehoppers (Membracidae: *Enchenopa binotata*). Gabriel D. McNett and Reginald B. Cocroft (Div. of Biological Sci., Univ. of Missouri, Columbia, 105 Tucker Hall, Columbia, MO 65211, gdmgw3@mizzou.edu)

Shifts to novel host plants can have dramatic consequences for a wide range of traits in plant-feeding insects. If the traits affected are mating signals, host shifts can provide a direct source of reproductive isolation. Mating signal evolution will be affected when changes in host use, either by use of a different plant species or plant part, lead to communication in a different signal environment. The sensory drive hypothesis predicts that signals should adapt to transmit efficiently in their local environment. Signal divergence, therefore, can occur where closely related insect species occur on host plants with different signal transmission properties. These predictions were tested in two closely related species in the Enchenopa binotata species complex (Hemiptera: Membracidae), host-specific plant-feeding insects that communicate using plant-borne vibrations. Their mating signals are relatively pure tones that vary among species in frequency (pitch), the most important signal trait for mate recognition. As predicted by sensory drive, it is shown that two closely related *E. binotata* species have evolved signals that transmit most efficiently in their contrasting communication environments. Changes in host use thus favor divergence of the signal trait most important for behavioral isolation. [Work supported by NSF.]

<center>2:30</center>

2pAB4. Strategies for seismic signal communication in spiders. Damian O. Elias, Andrew C. Mason (Dept. of Life Sci., Univ. of Toronto at Scarborough, 1265 Military Trail, Scarborough, ON M1C 1A4, Canada, elias@utsc.utoronto.ca), and Eileen A. Hebets (Univ. of Nebraska, Lincoln, NE 68588)

Communication is often hypothesized to be optimally designed for its specific signaling environment. While empirical studies have demonstrated this in several systems, the effect of high channel availability and heterogeneity has not been explored. Seismic (vibratory) communication presents a potential sender with a variety of potentially distinct signaling channels. Using examples from two types of spiders (*Salticidae* and *Lycosidae*), it is demonstrated that small animals use one of two strategies. Animals can be either: (1) specialized to the properties of a particular subset of signaling channels (narrow band signals) or (2) general to all signaling channels (broadband signals). Tradeoffs are discussed as well as the mate choice patterns associated with these different strategies. [Funding was provided by the National Science Foundation, the Natural Sciences and Engineering Council, and a Sigma Xi Society Grant-in-Aid.]

<center>**3:00-3:15 Break**</center>

<center>3:15</center>

2pAB5. Vibrational information in two life stages of the red-eyed treefrog: Agonistic communication signals and predation risk cues in an arboreal environment. Michael S. Caldwell, Karen M. Warkentin (Dept. of Biol., Boston Univ., 5 Cummington St., Boston, MA 02215), and J. Gregory McDaniel (Boston Univ., Boston, MA 02215)

Red-eyed treefrogs form dense mating aggregations and lay eggs in vegetation over neotropical ponds. Seismic information serves two important behavioral roles for this species. Adult males communicate with seismic signals during agonistic interactions, and embryos detect predators using vibrational cues. Males defending calling sites rapidly extend and contract their hindlimbs, shaking their bodies and the plant in tremulatory displays. This generates strong stereotyped substrate vibrations (12 ± 0.4 Hz, constant amplitude) that propagate to other males. Temporal, amplitude, and frequency properties of this signal are all distinct from common background vibrations and, in videotaped interactions, tremulation vibrations appear behaviorally relevant. In contrast, embryos experience predator and benign-source vibrations that overlap in temporal, amplitude, and frequency properties. These vibrations first pass through the gelatinous egg clutch and are shaped by its physical properties. Modal analysis indicates that natural free vibration frequencies of clutches are low (fundamental: 17 ± 2 Hz); these dominate both benign and dangerous direct disturbances. Additionally, the clutch acts as a frequency filter, rapidly attenuating energy over 200 Hz. Comparing the red-eyed treefrogs use of seismic information in communication and predator detection reveals the different requirements for perceptual strategies employing stereotyped and non-stereotyped information, and furthers our understanding of the seismic modality.

<center>3:45</center>

2pAB6. Vibrational risk assessment as a signal detection problem: Escape hatching of red-eyed treefrog eggs. Karen M. Warkentin, Michael S. Caldwell (Dept. of Biol., Boston Univ., 5 Cummington St., Boston, MA 02215, kwarken@bu.edu), and J. Gregory McDaniel (Boston Univ., Boston, MA, 02215)

The properties of cues from predators often overlap with background stimuli. Thus, prey may make two errors when assessing risk. They may miss cues and fail to defend themselves, or respond unnecessarily to false alarms. Although the incidence of these errors trades off, total error rates can be reduced only by adding information, either through increased sampling of one property or by sampling more cue properties. Adding cue properties likely increases processing requirements, and sampling predator cues entails risk. We examined a vibration-cued defense of the arboreal embryos of red-eyed treefrogs. These embryos use vibrations in snake attacks to cue behaviorally-mediated premature hatching and escape, but vibrations from benign sources rarely induce premature hatching. Missed cues and false alarms are costly; embryos that fail to hatch are eaten, and hatching prematurely increases predation by aquatic predators. Vibration playback experiments indicate that embryos attend to at least four nonredundant properties of vibrations: duration, spacing, stimulatory low frequencies, and inhibitory higher frequencies. They also adjust the amount of information they sample prior to hatching based on the time/risk cost of gathering it. This complexity of the risk assessment mechanism is consistent with strong selection against both missed cues and false alarms.

<center>4:15</center>

2pAB7. The vibration sense in large mammals and its role in communication: Elephants as a case study. Caitlin E. O'Connell-Rodwell (Dept. of Otolaryngol., Head & Neck Surgery, Stanford Univ., Stanford, CA 94305)

All mammals have the ability to detect vibrations and there are some notable cases of small mammals dedicating much of their sensory world to vibration detection such as the star nosed mole, the golden mole, the blind mole rat. In large mammals, the concept of using vibrations as a form of prey detection, predator avoidance, or communication has not been explored to a great extent. A few cases of these three uses of vibrations will be reviewed as they pertain to and may benefit the lion, kangaroo, and elephant seal, respectively. Elephants are well suited to communicate seismically, given their high amplitude, low frequency vocalizations that couple with and propagate in the ground as seismic signals. A unique combination of anatomical structures found on the elephant

would also facilitate seismic detection through either a bone conducted or somatosensory pathway, or both. A series of studies will be reviewed demonstrating the elephants ability to generate and propagate seismic energy from vocalizations and footfalls as well as their ability to detect and discriminate not only biologically meaningful seismic cues from noise, but also subtle differences between seismic vocalizations given in the same context.

4:45

2pAB8. Methods for studying seismic communication in elephants and other large mammals: Arrays, census techniques and more. Jason D. Wood, Caitlin E. O'Connell-Rodwell, Sunil Puria (Stanford Univ., Dept. of Otolaryngol.—Head and Neck Surgery, 801 Welch Rd., Stanford, CA 94305-5739), and Simon L. Klemperer (Stanford Univ., Stanford, CA 94305-2215)

Studying seismic communication poses unique methodological challenges. Methods for producing and monitoring seismic signals will be presented in the context of elephant seismic playback studies as well as large mammal footfall recordings for the purpose of censusing populations remotely. The use of multiple sensors in an array allows for a number of analyses to be conducted to determine the propagation of the signal. Specifically, arrays allow for the measurement of wave velocity and attenuation as well as determination of the mode of propagation. In addition, time of arrival differences can also be used to estimate the source of the signal which has important implications for measuring signal attenuation and in using seismic cues to estimate the number of animals in a group. One challenge in seismic playback studies is to ensure that there is no acoustic coupling of the seismic signal playback, the presence of which would make it impossible to rule out that any behavioral reactions of the study species are due to the acoustic signal that has leaked from the substrate. Matched filtering techniques have been used to detect any playback signal in the acoustic recording of the playback.

TUESDAY AFTERNOON, 5 JUNE 2007 SEMINAR THEATER, 1:00 TO 3:15 P.M.

Session 2pBBa

Biomedical Ultrasound/Bioresponse to Vibration: Therapeutic Ultrasound and Bioeffect

John S. Allen, III, Chair

Univ. of Hawaii, Dept. of Mechanical Engineering, Holmes Hall Room 302, 2540 Dole St., Honolulu, HI 96822

Contributed Papers

1:00

2pBBa1. Bubble proliferation in shock wave lithotripsy. Yuri A. Pishchalnikov, James A. McAteer (Dept. of Anatomy and Cell Biol., School of Medicine, Indiana Univ., 635 Barnhill Dr., Indianapolis, IN 46202-5120, yura@anatomy.iupui.edu), Michael R. Bailey (Univ. of Washington, Seattle, WA 98105-6698), James C. Williams, Jr. (Indiana Univ., Indianapolis, IN 46202-5120), and Oleg A. Sapozhnikov (Moscow State Univ., Moscow 119992, Russia)

Stone breakage is less efficient when lithotripter shock waves (SWs) are delivered at 2 Hz compared to slower 0.5–1-Hz pulse repetition rates (PRFs). This correlates with increased number of transient cavitation bubbles observed along the SW path at fast PRF. The dynamics of this bubble proliferation throughout the bubble lifecycle is investigated in this report. Cavitation bubbles were studied in the free-field of a shock wave lithotripter using fine temporal and microscopic spatial resolution (high-speed camera Imacon-200). A typical cavitation bubble became visible (radius > 10 μm) under the tensile phase of the lithotripter pulse, and at its first inertial collapse emitted a secondary SW and formed a micro-jet, which then could break up forming ~ 25 micro-bubbles. Subsequent rebound and collapse of the parent bubble appeared to produce a further 40–120 daughter bubbles visible following the rebound. Preexisting bubbles hit by the lithotripter SW also formed micro-jets and broke up into micro-bubbles that grew and coalesced, producing irregular-shaped bubbles that, in turn, broke into micro-bubbles upon subsequent inertial collapse. A conventional NTSC-rate camcorder was used to track cavita-

tion bubbles from pulse-to-pulse, showing that a single bubble can give rise to a cavitation cloud verifying high-speed video results. [Work supported by NIH-DK43881.]

1:15

2pBBa2. Blood mimicking fluid for high-intensity focused ultrasound applications. Yunbo Liu, Subha Maruvada, Keith A. Wear, Bruce A. Herman (Food and Drug Administration, Ctr. for Devices and Radiological Health, 9200 Corporate Blvd., M.S. HFZ-170, Rockville, MD 20850), and Randy L. King (Stanford Univ., Stanford, CA 94305)

A blood-mimicking fluid (BMF) having viscosity, attenuation, and backscatter similar to that of human blood has been developed for the acoustic and thermal characterization of HIFU ablation devices. The BMF consists of gellan gum, low density polyethelene particles, and nylon particles dispersed in water, glycerin, and alcohol. The BMF was characterized for attenuation coefficient, speed of sound, backscatter coefficient, viscosity, thermal diffusivity, and thermal conductivity. The low attenuation as well as the sound speed and the bloodlike backscatter indicate the usefulness of the BMF for ultrasound imaging and flow applications. These properties, along with thermal conductivity and diffusivity also within the range found in human tissue and the ability to withstand temperature increases above 70 °C with no significant change in its properties, make this material appropriate for HIFU applications. This fluid also ex-

hibits viscosity varying as a function of velocity (or shear rate).

1:30

2pBBa3. Formation of shock waveforms and millisecond boiling in an attenuative tissue phantom due to high-intensity focused ultrasound.
Michael S. Canney, Michael R. Bailey, Vera A. Khokhlova, and Lawrence A. Crum   (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

Nonlinear propagation effects during high-intensity focused ultrasound (HIFU) treatments can induce shocks in the acoustic waveform, dramatically accelerate heating rates, and result in rapid boiling of tissue at the focus. Localized boiling can be used for targeting and calibration of clinical HIFU treatments. In our previous work, millimeter size boiling bubbles were observed in several milliseconds in a weakly absorptive transparent tissue phantom, and temperature rise to 100 °C was calculated using weak shock theory from experimentally measured and numerically simulated focal waveforms [Canney *et al.*, J. Acoust. Soc. Am. **120**, 3110 (2006)]. In this work, experiments are extended to an opaque phantom that has higher attenuation (0.5 dB/cm/MHz in the new phantom versus 0.15 dB/cm/MHz in the previous one) more similar to real tissue. Focal acoustic waveforms are measured using a fiber optic probe hydrophone and time to boil is monitored using a 20-MHz acoustic detector. Modeling of experimental conditions is performed with a KZK-type numerical model. Results demonstrate that although higher source amplitude is needed to attain the same focal amplitudes in the new, more attenuative phantom, similar amplitude shocks can be formed, resulting in equally fast heating rates. [Work supported by NIH DK43881 and NSBRI SMS00402.]

1:45

2pBBa4. Parametric study of a shape based inversion for detecting high-intensity focused ultrasound lesions.   Bruno Durning   (Elec. and Comput. Eng., Northeastern Univ., Boston MA 02115), Robin O. Cleveland   (Boston Univ., Boston MA, 02215), and Eric L. Miller   (Tufts Univ., Medford, MA 02155)

A key problem in the practical use of high intensity focused ultrasound (HIFU) as a tool for cancer treatment is the non-invasive characterization of the regions of tissue that have successfully been necrosed. Previously, we proposed an approach to image guidance, based on the use of RF data obtained from a diagnostic ultrasound transducer and a shape-based inverse scattering approach. Specifically, it was assumed that the lesion has an ellipsoidal shape defined by its center, size, orientation, and contrast (in sound-speed and attenuation) compared to the background. An inverse-type method was used to identify the ellipsoid parameters from the RF data. In this work we explore the robustness of this approach to a variety of conditions likely to be encountered in practice, specifically, the presence of an aberratting layer in the path, the formation of non-ellipsoidal lesions, for example, a "tadpole" shaped lesion that is commonly formed during HIFU, and the presence of multiple objects. Experiments using a clinical scanner and tissue phantoms are reported and we evaluate the method's efficiency to different shapes and number of objects. [Work supported by NIH and CenSSIS.]

2:00–2:15   Break

2:15

2pBBa5. Effect of intermittent applications of continuous ultrasound on the viability, proliferation, morphology and matrix production of chondrocytes in 3-D matrices.   Joseph A. Turner   (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, Lincoln, NE 68588, jaturner@unl.edu), Sandra Noriega, Tarlan Mammedov, and Anu Subramanian   (Univ. of Nebraska-Lincoln, Lincoln, NE 68588)

Chondrocytes, the cellular component of articular cartilage, have long been recognized as strain-sensitive cells, have the ability to sense mechanical stimulation through surface receptors and intracellular signaling pathways. This response has been exploited to facilitate chondrocyte culture in *in vitro* systems such as those that use hydrostatic pressure, dynamic compression, hydrodynamic shear, as well as low-intensity pulsed ultrasound (US). While the ability of US to influence chondrogenesis has been documented, the precise mechanisms of US induced stimulation are unclear. Thus, a critical need remains to evaluate the impact of US on chondrocytes in 3-D cultures, a necessary microenvironment for maintaining chondrocyte phenotype. Here, chondrocytes seeded in 3-D scaffolds were subjected to continuous ultrasound stimulation at several frequencies for the same number of cycles, applied twice in a 24-hour period. Non-US stimulated scaffolds served as the control. Both groups were maintained in culture for 10 days and were assayed at the conclusion of the culture period (total DNA content, morphology, and cartilage specific gene expression). Our results show that chondrocytes stimulated with continuous US for predetermined time intervals possessed 1.2 to 1.4-times higher cellular viability than the control as well as higher levels of type-II collagen and aggrecan mRNA expression.

2:30

2pBBa6. Lung hemorrhage at and near resonance: Pulse duration and pulse number.   Diane Dalecki, Sally Z. Child, and Carol H. Raeman (Dept. of Biomed. Eng., and the Rochester Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627)

Research from our lab has shown that low-frequency ($\sim$100–1000 Hz) underwater sound can produce mammalian lung hemorrhage. For adult mice, the pressure threshold for lung damage is $\sim$2 kPa for a 3-min continuous wave exposure at the lung resonance frequency ($\sim$300 Hz). Sound-induced lung hemorrhage was studied for exposure below (200 Hz) and above (500 Hz) lung resonance. The threshold for lung damage is lowest for exposure at the lung resonance frequency. Furthermore, a series of experiments was performed to investigate effects of pulse duration and number of pulses on lung hemorrhage produced by exposure to underwater sound at lung resonance frequency. The resonance frequency of each mouse lung was determined using an acoustic scattering technique. In one set of experiments, the extent of lung hemorrhage was assessed for increasing pulse durations ranging from 1–180 s. In another set of experiments, extent of lung hemorrhage was assessed for increasing number of pulses. Extent of lung hemorrhage increased with increasing pulse duration and increasing number of pulses. The lung resonance frequency can shift higher or lower during sound exposure. The results of this work are relevant to establishing safety guidelines for swimmers and divers exposed to underwater sound fields.

2:45

2pBBa7. Effective medium model of human lung response to low-frequency sound.   Mark S. Wochner, Yurii A. Ilinskii, Mark F. Hamilton, and Evgenia A. Zabolotskaya   (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, mwochner@arlut.utexas.edu)

This presentation is an extension of work described previously [Ilinskii *et al.*, J. Acoust. Soc. Am. **120**, 3194 (2006)] on modeling the response of human lung to low-frequency underwater sound. A lumped element model with alveoli represented as truncated octahedra forming a periodic lattice with cubic symmetry was developed to capture the microscopic properties of collagen and elastin. The lattice is deformed quasistatically to determine the three elastic constants associated with the macroscopic behavior of the system. In reality lung tissue is likely isotropic and therefore a method of averaging is utilized to determine the two Lamé constants of the effective medium. The volume of the lung is varied to simulate tidal breathing. The Lamé constants are determined for the given lung volume and used in a commercial finite element package to calculate the amplitude of vibration due to low-frequency acoustic excitation. The resulting spectral response and scattered sound field are calculated for a water-loaded viscoelastic sphere composed of the effective lung medium. Increase in lung volume tends to decrease the resonance frequency, but increasing stiffness due to collagen tends to increase the resonance frequency. Competition of these effects is discussed. [Work supported by ONR and ARL IR&D.]

3:00

2pBBa8. Time reversal acoustic focusing with liquid-filled reverberator. Yegor Sinelniov, Andrey Vedernikov (ProRhythm Inc., 105 Comac St., Ronkonkoma, NY 11779), Alexander Sutin, and Armen Sarvazyan (Artann Labs., Inc., West Trenton, NJ 08618)

The use of solid acoustic reverberators for temporal and spatial focusing of acoustic energy based on time reversal acoustic (TRA) principle was described in the literature. We experimentally and theoretically demonstrated that the liquid-filled TRA reverberators can also efficiently focus acoustic energy and compared the advantages of solid and water-filled reververators. A theoretical model was developed to explore the factors defining the efficiency of TRA focusing, providing the design constraint for the water-filled reververators. The model allows evaluation of the spatial and temporal characteristics of TRA focused acoustic field formed as a result of numerous reflections of ultrasound within the reverberator. The TRA focused field structure, its dependence on the shape and size of the resonator, ultrasound frequency, and attenuation were investigated for the standard and binary modes of TRA focusing. The experiments with custom built liquid-filled acoustic resonators confirmed the predictions of a developed theory, including sharpening of the focal area and increasing of the focal intensity in a binary mode of TRA focusing. The conducted research supports suggestion that such liquid-filled reverberators can be considered as a base for a miniature limited range ablation devices for biomedical applications.

TUESDAY AFTERNOON, 5 JUNE 2007　　　　　　　SEMINAR THEATER, 3:30 TO 5:00 P.M.

Session 2pBBb

Biomedical Ultrasound/Bioresponse to Vibration: Imaging and Detection Theory

Bruno Durning, Chair
Tufts Univ., Dept. of ECE, Medford, MA 02155

Contributed Papers

3:30

2pBBb1. Fast computation of spatial transfer function for ultrasound imaging. Emre H. Guven (Dept. ECE, Northeastern Univ., Boston, MA 02115, eguven@ece.neu.edu), Eric Miller (Tufts Univ., Medford, MA 02155), and Robin Cleveland (Boston Univ., Boston, MA 02215)

A fast method for computing the spatial transfer function of ultrasound transducers is presented with application to rectangular transducers that are focused in one axis only. No closed form solutions exist for this application but several numerical techniques have been described in the ultrasound imaging literature. Our motivation is the rapid calculation of imaging kernels for physics based diagnostic imaging where current methods are computationally demanding. The spatial transfer function to be calculated is a spatial convolution of the transducer surface and the Green's function. A 3-D version of the overlap-save method, which is a method for digital filtering of long data sequences, has been employed to obtain a fast computational algorithm. Further efficiency is gained by using separable approximations of the 3-D convolution through the singular value decomposition. The tradeoff between accuracy and spatial sampling rate is explored to determine appropriate parameters for a specific transducer. Comparisons with standard tools such as Field II will be presented. [Work supported by NIH and CenSSIS.]

3:45

2pBBb2. The causal impulse response for circular pistons in viscous media. James F. Kelly and Robert J. McGough (Dept. of Elec. and Computer Eng., Michigan State Univ., East Lansing, MI 48824)

A causal impulse response for the Stokes wave equation is derived for calculations of transient pressure fields generated by circular pistons in viscous media. The causal Green's function is numerically verified using the material impulse response function approach. The causal, lossy impulse response for a baffled circular piston is then calculated within the near-field and the far-field regions using expressions previously derived for the fast near-field method. Expressions for apodized pistons evaluated in the far-field region are also demonstrated. Transient pressure fields in viscous media are computed with the causal, lossy impulse response and compared to results obtained with the lossless impulse response. The numerical error in the computed pressure field is quantitatively analyzed for a range of viscous relaxation times and piston radii. Results show that the largest errors are generated in locations near the piston face and for large relaxation times, and errors are relatively small otherwise. These results suggest that this causal impulse response is ideal for time-domain calculations that simultaneously account for diffraction and frequency-dependent attenuation in viscous media. [This work was partially supported by NIH Grant 1R01 CA093669.]

4:00

2pBBb3. Stress and strain relaxation mechanisms in soft tissue: An inverse problem and solution. Ricardo Leiderman (Program of Mech. Eng.—EE/COPPE, Federal Univ. of Rio de Janeiro, Rio de Janeiro, Brazil), Gearoid P. Berry, Jeffrey C. Bamber (Inst. of Cancer Res. and Royal Marsden Hospital NHS Trust, Sutton, UK), Assad A. Oberai (Rensselaer Polytechnic Inst., Troy, NY), and Paul E. Barbone (Boston Univ., Boston, MA)

Elastography refers to a collection of ultrasound imaging techniques that allow mechanical strain distributions to be imaged and noninvasively quantified *in vivo*. The time scales over which the tissue response is typically measured range from about a millisecond (the typical duration of a radiation force "push pulse") to about one second (the typical time scale of freehand quasistatic compression) used in strain imaging. Soft tissue is widely recognized as having both fluid and solid phases which can move independently of each other, giving rise to stress/strain relaxation effects. Furthermore, the fluid exists within several "compartments" of the soft tissue, notably, the vasculature (including both the hemal and lymphatic vessels) and the extravascular space (i.e., the interstitium). Of course, due to permeability of microvessel walls in both vascular networks, fluid is often exchanged between these compartments. Through mathematical models of tissue deformation, we explore the macroscopic effects of different microscopic relaxation mechanisms. We show that different microscopic relaxation mechanisms result in different spatio-temporal patterns of strain relaxation. We then focus on one such mechanism, and we formulate and solve an inverse problem that measures the microvessel density and permeability from the time-history of a measured strain distribution.

4:15

2pBBb4. Is the Kramers-Kronig causal relationship between ultrasonic attenuation and dispersion maintained when phase aberrations distort the field incident on a phase sensitive aperture? Adam Q. Bauer, Karen R. Marutyan, Mark R. Holland, and James G. Miller (Washington Univ. in St. Louis, One Brookings Dr., Campus Box 1105, St. Louis, MO 63130)

The objective of this investigation was to determine whether the causality-induced link between ultrasonic attenuation and dispersion remains valid when the measured signal loss arises as a consequence of phase cancellation at the face of a phase-sensitive receiver rather than from intrinsic losses within the medium under study. The frequency-dependent apparent attenuation and phase velocity were obtained from through-transmission measurements of two flat and parallel plastic polymer plates, Plexiglas^{TM} and Lexan^{TM}, exhibiting an approximately linear with frequency attenuation coefficient and logarithmic with frequency phase velocity. Phase distortion was achieved by machining a step into one side of each plastic plate. Through-transmission measurements were performed in a water tank using 5 MHz center frequency single element planar transmitting and receiving transducers. The causal (Kramers-Kronig) link between apparent phase velocity and the apparent attenuation coefficient in the presence of phase distortion was examined for both plastics over a bandwidth ranging from 3 to 7 MHz. Results demonstrate that the Kramers-Kronig link between the apparent attenuation coefficient and apparent phase velocity dispersion remains causally consistent even in the presence of aberrating media.

4:30

2pBBb5. Three dimensional ultrasound elasticity imaging. Michael S. Richards, Jonathan M. Rubin (Univ. of Michigan, University Hospital, Dept. of Radiology, Ann Arbor, MI 48109), Assad A. Oberai (Rensselaer Polytechnic Inst., Troy, NY), and Paul E. Barbone (Boston Univ., Boston, MA 02115)

The aim of our work is to develop and evaluate an ultrasound (US) technique to quantitatively measure and image the shear modulus of soft tissues in three dimensions (3D). It is widely recognized that breast tissue pathologies, such as neoplasia, often alter biomechanical properties. Thus,

the intended application of our work is the detection and characterization of breast tissue lesions, via elastic modulus imaging, to improve the specificity of breast cancer screening. To that end, we have designed and characterized algorithms which 1) provide 3-D motion estimates from 3-D US images and 2) solve the 3-D inverse problem to recover shear elastic modulus. The size and contrast accuracy of the reconstructed modulus distributions in tissue mimicking phantoms are presented. Inclusions as small as 5mm, some with contrasts marginally above unity, were successfully and clearly reconstructed. The effect of the boundary conditions and regularization methods on the reconstructed modulus images and the uniqueness of the solution are also discussed. In addition, preliminary modulus reconstructions created from clinical 3-D US breast images acquired in a mammography mimicking system are presented.

4:45

2pBBb6. Application of the continuous wavelet transform to acoustic emissions generated by joint and muscle motion. Wayne Fischer and Joe Guarino (Dept. of Mech. Eng., Boise State Univ., 1910 University Dr., Boise, ID 83725, wfische@boisestate.edu)

The subject matter of this presentation will refer to the preliminary development of a non-destructive and non-invasive diagnostic method using low-frequency acoustic emissions to characterize and diagnose joint and muscle disorders. Low-frequency acoustic emissions have had limited effectiveness when used in the past to characterize and diagnose joint and muscle disorders because Fourier transform techniques were not capable of describing when a specific frequency occurred during an acoustic emission event. With the development of techniques that decompose a signal into time-frequency representations using wavelet analysis it is now possible to describe complex transient signals. While the continuous wavelet transform (CWT) method has already made an impact within many areas of the medical sciences such as EEG analysis, ECG analysis, and DNA-sequence analysis, the method is now becoming popular for analyzing muscle and joint acoustic emissions from the hip, knee, and elbow joints. This paper will present the results of our application of the CWT to the analysis of acoustic emissions from the shoulder joint.

TUESDAY AFTERNOON, 5 JUNE 2007　　　　　　　　　　　　　　　　SALON B, 1:30 TO 4:15 P.M.

Session 2pEA

Engineering Acoustics, Underwater Acoustics and Signal Processing in Acoustics: Acoustic Technologies for Coastal Surveillance and Harbor Defense

Roger T. Richards, Chair
Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841

Chair's Introduction—1:30

Invited Paper

1:35

2pEA1. Swimmer detection sonar network. Peter J. Stein, Amy Vandiver, (Sci. Solutions, Inc., 99 Perimeter Rd., Nashua, NH 03063, pstein@scisol.com), and Geoffrey S. Edelson (BAE Systems, Nashua, NH 03061)

The Swimmer Detection Sonar Network (SDSN) was designed to be a cost effective method to protect a large area from terrorist swimmers. The in-water system consists of a number of sonar nodes, each consisting of a set of air-backed parabolic dish transducers that are used to both transmit and receive in sets of independent narrow beams. Each dish/channel can transmit and process its own unique signal. The Sonar Network is therefore composed of a distributed group of nodes each providing a set of independent narrow-beam sonars. This allows for strong rejection of interfering clutter, resulting in excellent performance. The nodes/channels are also synchronized, allowing for multi-static signal processing. A suite of signal processing algorithms has been developed for

automated detection, tracking, and classification, including a fine-bearing estimation that achieves approximately 1 deg accuracy. These algorithms will be discussed. Data will be shown that shows the performance against swimmers and the possibilities for bi-static detection and classification. The SDSN has been productized and is now available for wide-scale deployment. It is a cost effective means for large harbor facilities to reliably detect underwater threats at long ranges.

Contributed Papers

2:00

2pEA2. Clearance of harbor navigation channels with high-resolution bottom surveys at 10 knots. James Glynn, Jr, Christian de Moustier, and Lloyd Huff (Ctr. for Coastal and Ocean Mapping, UNH, 24 Colovos Rd., Durham, NH 03824)

The safety of navigation channels leading to or within a harbor can be ensured by repeat high-resolution bottom surveys to detect new hazards or obstacles. This task requires a sonar system capable of acquiring high-resolution acoustic backscatter imagery and bathymetry. Because of the generally shallow water depths (20 m), survey efficiency dictates data collection over wide swaths at high speeds. We have demonstrated these capabilities in October 2006 in New York harbor, NY, with a Klein 5410 multibeam side-looking sonar system yiedling swath coverage of at least 7 times the water depth at a nominal survey speed of 10 knots. This has been achieved by integrating the hull-mounted sonar system with precision attitude (0.01 deg) and timing (10-6 s) references, and by designing and implementing a full vector processing algorithm to obtain co-registered high-resolution bathymetry and acoustic backscatter imagery, allowing us to resolve sub-meter targets. [Work funded by NOAA-NMFS].

2:15

2pEA3. Detection and target echo enhancement in shallow water using the Decomposition of the Time Reversal Operator. Claire Prada, Julien de Rosny, Dominique Clorennec, Jean-Gabriel Minonzio, Alexandre Aubry, Mathias Fink (Lab. Ondes et Acoustique, ESPCI, Paris, France), Lothar Bernière, Sidonie Hibral, Philippe Billand, and Thomas Folégot[a)] (Altran Technologies, Technopôle Brest Iroise, Brest, France)

A rigid 24-element source-receiver array (SRA) in the 10 to 15 kHz frequency band, connected to a programmable electronic system was deployed in the bay of Brest (France) during spring 2005. In this 10 to 18 m deep environment, backscattered data from submerged targets were recorded. Successful detection and focusing experiments in very shallow water using the Decomposition of the Time Reversal Operator (DORT method) are shown. The ability of the DORT method to separate the echo of a target from reverberation is demonstrated. For a target close to the bottom at 350 m range, active focusing using the first invariant of the time reversal operator was achieved, showing a target echo enhancement. Furthermore, the localization of the target in the water column is obtained using RAM [Collins *et al.*, J. Acoust. Soc. Am. **89** 3 (1991)]. [Work supported by the French DGA/SPN under contract number 02 77 154 00 470 75 53.] [a)] now affiliated to NATO Undersea Research Centre, La Spezia, Italy

2:30

2pEA4. Estimation of passive acoustic threat detection distances in estuarine environments. Brian Borowski, Heui-Seol Roh, Barry Bunin, and Alexander Sutin (Stevens Inst. of Technol., 711 Hudson St., Hoboken, NJ 07030)

The Maritime Secure Laboratory (MSL) at Stevens Institute of Technology supports research in a range of areas relevant to harbor security, including passive acoustic detection of underwater threats. The difficulties in using passive detection in an urban estuarine environment include intensive and highly irregular ambient noise and the complexity of sound propagation in shallow water. MSL measured the main parameters defining the detection distance of a threat: source level of a scuba diver, transmission loss of acoustic signals, and ambient noise. The source level of the diver was measured by comparing the divers sound with a reference signal from a calibrated emitter placed on his path. Transmission loss was measured using the transmission of a sweep signal (1–100 kHz) from the calibrated emitter. The passive sonar equation was then applied to estimate

the range of detection. Estimations were done for various recorded noise levels, demonstrating how fluctuations in noise level and the mobility of the diver influence the effective range of detection. Finally, analytic estimates of how a hydrophone array improves upon the detection distance calculated by a single hydrophone are shown. [This work was supported by ONR project No. N00014-05-1-0632: Navy Force Protection Technology Assessment Project.]

2:45

2pEA5. Acoustic noise produced by ship traffic in the Hudson River estuary. Heui-Seol Roh, Barry Bunin, George Kamberov, and Alexander Sutin (Stevens Inst. of Technol., 711 Hudson St., Hoboken, NJ 07030)

This paper presents results of measurements of acoustic noise in Hudson River Estuary near Manhattan in the frequency band 10–100 kHz. The Estuary has very complex sound propagation conditions due to the extremely shallow and high time- and space variability of the water characteristics. The acoustic noise was recorded by a set of hydrophones and the acoustic measurements were accompanied by ship traffic video recording using the video-based Surface Traffic Surveillance system. This video system allowed us to map various boats and ships and to find distances between them and the hydrophone system. The measurements provided acoustic noise data for different kinds of ships in Hudson River, their dependencies on frequencies, and distances. The measurements of noise for various distances were applied for estimation of sound attenuation in a wide frequency band. We calculated the sound attenuation coefficient showing the attenuation loss of an acoustic signal in addition to cylindrically spreading loss. The recorded levels of acoustic noise and estimation of transmission loss were used for estimations of detection distances of underwater threats that are presented in our other paper. [This work was supported by ONR Project No. N00014-05-1-0632: Navy Force Protection Technology Assessment Project.]

3:00–3:15   Break

3:15

2pEA6. Performance of a high-frequency acoustic forward-scatter barrier in a dynamic coastal environment. Karim G. Sabra, Stephane Conti, Philippe Roux, W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., UC San Diego, San Diego, CA 92093-0238, ksabra@mpl.ucsd.edu), J. Mark Stevenson, Alessandra Tesei, Piero Guerrini, Piero Bonni (NATO Undersea Res. Ctr., La Spezia, Italy), and Tuncay Akal (UC San Diego, San Diego, CA 92093-0238)

This paper investigates the performance of a high-frequency ''barrier'' or ''tripwire'' surveillance system based on forward scattering for shallow water (<20 m depth) choke points, such as harbor entrance. Harbor entrances are usually regions of very high ambient noise and often are reverberation-limited environments, making them challenging for some traditional ASW techniques, i.e., techniques inherited from blue-water ASW. Detecting a target by measuring its forward-scattered field is of interest because sound scattered in the forward direction is generally higher than in the backward direction. We will present a series of proof-of-concept experiments to test the feasibility of an acoustic tripwire based on forward scattering using two vertical source and receive arrays operated over a 135-m range at a center frequency of 15 kHz with about 8 kHz of bandwidth. The signal of interest was the aberration (in space and time) caused by the acoustic forward scattering field of a crossing object (scuba tank, 2-m-long steel cylinder and 1-m-diam steel sphere). Autonomous signal processing techniques based on the spatio-temporal coherence of

the signals measured across the vertical receive array were applied to enhance the signal of interest, yielding a usable detection threshold of several dB.

3:30

2pEA7. Improved range and characterization of harbor security systems through bistatic detection. Marcia Isakson, Colin Bown, Karl Rehn, and Nathan Crowe (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

There are three critical design drivers for harbor security systems: cost per unit area of coverage, coverage of geometrically denied areas such as very shallow waters and under and around piers, and probability of false alarm. Bistatic detection can address all three of these issues. First, many harbor security systems operate with overlapping coverage. These adjacent systems offer an existing source of bistatic signals. By analyzing the signal from these adjacent sources, the coverage for each sound head is increased with no increase in system cost. Second, the strategic placement of additional sources will allow coverage in denied areas for little additional cost. Lastly, additional detection will lower false alarm rates. In this study, measurements from an overlapping harbor security system will be analyzed to determine the effect of additional bistatic analysis on detection thresholds and coverage range.

3:45

2pEA8. Application of maximum length sequences to photoacoustic chemical analysis. Ralph T. Muehleisen and Arash Soleimani (Civil and Architectural Eng., Illinois Inst. of Technol., Chicago, IL 60616, muehleisen@iit.edu)

There is a great need for inexpensive, rugged, portable, and versatile chemical detectors for use in both security and environmental measurement. Current sensors that meet the inexpensive, portable, and rugged criterion are usually limited in sensitivity and dedicated to detecting a limited number of chemicals. Sensors designed for airborne measurement typically cannot be used to measure surface chemicals. The photoacoustic method is one of the most sensitive techniques for trace analysis in both the air and on surfaces. Unfortunately, photoacoustic systems are typically expensive, usually not portable, and usually cannot be used for both airborne and surface chemical analysis. A new design for a photoacoustic measurement system is proposed that utilizes maximum length sequences (MLS) to modulate multiple, inexpensive, fixed frequency, laser diodes. The MLS modulation and post-acquisition correlation processing should provide a very high signal-to-noise measurement without need for lock-in amplifiers, high power pulsed lasers, optical modulators, or resonant measurement cells. These changes should allow for the design of a rugged, portable, inexpensive photoacoustic measurement system that can be used for both airborne and surface trace chemical detection.

4:00

2pEA9. Acoustic emission mechanism from scuba diving equipment. Dimitri M. Donskoy (Davidson Lab., Stevens Inst. of Technol., 711 Hudson St., Hoboken, NJ 07030)

The energy-based analysis of the underwater acoustic emission mechanism from scuba diving equipment shows that the primary originating source of acoustic emission is turbulent air flow pressure fluctuations during the inhale phase of breathing. A scuba air tank contains a substantial amount of energy in the form of compressed air. This energy is gradually released as the air is consumed during breathing. As air expands from high pressure in the tank to lower pressure supplied to the second stage regulator, it rushes through the first stage regulator's valve and channels creating highly turbulent air flow. The turbulent pressure fluctuations excite structural vibrations of the valve housing and connected air tank and piping in a very broad frequency range spreading above 100 kHz. In turn, vibration causes sound emission into surrounding medium (water). [Work supported by ONR].

TUESDAY AFTERNOON, 5 JUNE 2007   ALPINE WEST, 1:30 TO 3:45 P.M.

Session 2pMUa

Musical Acoustics: Voice Production and Pedagogy

Ingo R. Titze, Chair
Univ. of Iowa, Speech Pathology and Audiology, Iowa City, IA 52242-1012

Chair's Introduction—1:30

Invited Papers

1:35

2pMUa1. Singers' self-perception of harmonics. Lisa Popeil (Voiceworks, 14431 Ventura Blvd. #402, Sherman Oaks, CA 91423)

Historically, singers have relied on the sensation of resonance as a guide in vocal production. The use of terms such as focus and placement have been primary tools in traditional vocal pedagogy. In a newer method of acoustic training, singers can be taught to discern three bands of harmonics which are key elements of vocal timbre: ring, brightness, and nasality. Each of these harmonic bands has particular sonic and expressive characteristics and are easily amplified or attenuated. Ring, at 2500–3500 Hz, is the most penetrating of the three and is acutely heard by the human ear. Ring aids singers in being heard without electronic amplication. Brightness, comprised of harmonics in the approximate range of 5–15 kHz, lacks the penetrating quality of ring but is an important conveyor of the emotion of happiness and expressive traits such as sincerity and innocence. Nasality, broadly occuring in the 200–2000 Hz range, has a characteristic buzziness and is used in commercial singing styles for amplitude and expressive enhancement. Live vocal demonstrations will show how a singer can create, isolate, and combine ring, brightness, and nasality.

1:55

2pMUa2. Acoustically-guided vocal tract modifications for singing. Brad Story   (Univ. of Arizona, P.O. Box 210071, Dept. of Speech, Lang., and Hearing Sci., Tucson, AZ 85721)

The sound quality of a specific vowel can be dramatically altered by subtle modifications of the vocal tract shape. These modifications create changes in the pattern of formant frequencies. For example, the well-known singing formant, which is a clustering of resonance frequencies, is typically the result of constricting the epilaryngeal space or expanding the lower pharyngeal space to create a large cross-sectional area discontinuity between them. Other modifications such as lip protrusion/spreading or larynx lowering/raising will also impose changes on the formant frequency pattern that may be desirable for singing or speech production. This presentation will focus on the acoustic sensitivity of the resonance frequencies to subtle perturbations of specific vowel configurations. Using calculated sensitivity functions, it will be shown how specific regions along the vocal tract can be constricted or expanded to perturb one or more of the formant frequencies. In effect, this technique provides a means of "tuning" the vocal tract shape to produce a desired frequency response. [Work supported by NIH R01-DC04789.]

2:15

2pMUa3. The 1:6 ratio in vocal pedagogy. Brian B. Monson   (Dept. of Speech, Lang., and Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721)

The so-called singer's formant or singer's formant cluster is created in the vocal tract by a narrow epilaryngeal tube opening, such that the ratio of the cross-sectional area of the pharynx is approximately six times greater than that of the epilarynx outlet. This appears to be accomplished by maintaining a comfortably low laryngeal position while manipulating other structures of the vocal tract. Experimental research was conducted to examine the potential of pedagogical implementation of this principle. Experimentation consisted of training graduate voice student subjects at Brigham Young University to use the one-to-six ratio to achieve the singer's formant. Using the acoustical output from the singers for analysis, predictions of actual vocal tract shapes before and after vocal tract manipulation will be shown. The results of this study may offer valuable information regarding the training of voice students to access the singer's formant.

2:35

2pMUa4. Semi-occluded vocal tract postures and their application in the singing voice studio. John P. Nix   (Music Dept., The Univ. of Texas at San Antonio, San Antonio, TX 78249) and C. Blake Simpson   (Univ. of Texas Health Sci. Ctr. at San Antonio, San Antonio, TX 78249)

Many singing teachers employ semi-occluded vocal tract postures in their teaching. These postures can be divided into three types: (1) postures where the semi-occlusion remains consistent over time, as in singing vowels into a straw, the sustained phonation of voiced fricative consonants and the sustained phonation of nasal consonants; (2) postures where the occlusion is oscillatory, as in the lip buzz, the raspberry, and the rolled /r/; (3) postures where the semi-occlusion or occlusion is very transitory, as is found in the semi-vowels /j/ and /w/ and the voiced plosive consonants. The presentation will review the benefits of these semi-occluded postures, show videoendoscopic images of the pharynx and larynx and electroglottography of the vocal folds while the semi-occluded postures listed above are used as pilots to the vowels /i/, /u/, and /a/, and will demonstrate some of the traditional uses of the semi-occluded postures in singing pedagogy.

2:55

2pMUa5. Source-vocal tract interaction in singing. Ingo R. Titze   (Dept. of Speech Pathol. & Audiol., Univ. of Iowa, Hawkins Dr., Iowa City, IA 52242)

It has been known for centuries that certain vowels are easier to sing on certain pitches. Much of traditional and modern singing pedagogy is focused on this vowel F0 interaction. Unlike most musical instruments, where the resonator is specifically designed to resonate many of the source frequencies at every note played, the vocal instrument relies on the use of inertive reactance (away from resonance) to boost various clusters of harmonics. Acoustic reactance is positive (inertive) below a vocal tract resonance (a formant), zero at resonance, and negative (compliant) above resonance. The skill in selecting a vowel for a given F0 is to place as many of the key harmonics (2F0, 3F0, and 4F0) on the inertive portions of the reactance curves. The process is aided by shrinking the epilarynx tube, which biases the reactance curves toward positive values over the entire F0 range. [Work supported by NIDCD.]

Contributed Papers

3:15

2pMUa6. A real-time display system for singing voice development.
David M. Howard, Helena Daffern, Jude Brereton (Audio Lab.,
Intelligent Systems Res. Group, Dept. of Electron., Univ. of York,
Helsington, York, YO10 5DD, UK), Graham F Welch, Evangelos
Himonides (Univ. of London, London, WC1H 0AL UK), and Andrew
W. Howard (SARAND Ltd., Cambridge, UK)

Central to any pedagogical nurturing of singing development is the
provision of some form of meaningful feedback to the developing singer.
Singing teachers draw on their personal experiences within an essentially
oral culture where expertise is handed down from teacher to student gen-
eration by generation. The teacher is engaged in a psychological transla-
tion of the student's performance, and a dual possibility thereby exists for
the misinterpretation of information; the teacher may not describe the
student's performance appropriately and/or the student may not understand
how to modify his or her singing behavior as intended. The provision of
quantitative visual feedback provides a physical basis for vocal develop-
ment. This paper will describe a real-time display system for singing train-
ing known as "WinSingad," which enables acoustic and voice source
aspects to be viewed in real-time on a PC. The system has been used in
singing lessons and has been universally welcomed by students and teach-
ers alike for its ease of use and display clarity. Specific aspects of voice
development are now linked with particular displays and these will be
explored in this presentation.

3:30

2pMUa7. A non spectrum-based visual display tool. Peter R.
Nordquist, R. Dean Ayers, Daisuke Kato, and Lewis Nakao (Southern
Oregon Univ., 1250 Siskiyou Blvd., Ashland, OR, 97520)

A software program was developed to help musicians learn to tune just
intonation intervals. A reference tone is played and Lissajous-like figures
present a visual representation of the interval the musician's sound makes
with the reference tone. This approach is interesting, because the musician
receives coordinated aural and visual feedback in real time, and this visual
representation is produced without the overhead of doing a Fourier trans-
form. Experiments that modify the figures to produce a usable representa-
tion of timbral differences, also using a non spectrum-based algorithm, are
planned. [Work supported by the Veneklasen Foundation.]

TUESDAY AFTERNOON, 5 JUNE 2007 ALPINE WEST, 4:15 TO 5:30 P.M.

Session 2pMUb

Musical Acoustics: Mini-Concert of Yodeling as a Style of Vocal Production

William J. Strong, Chair
Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602

Chair's Introduction—4:15

Kerry Christensen—yodeler—became interested in alpine yodeling while spending two years in Austria. He used his yodeling
skills at Walt Disney's Epcot Center for almost eight years and will visit over 20 states and two or three countries this year. Kerry
specializes in all styles of yodeling. He will perform alpine, classical, humorous, and other yodeling styles during this mini-concert.

TUESDAY AFTERNOON, 5 JUNE 2007 CANYON ROOM A, 1:30 TO 3:30 P.M.

Session 2pNS

Noise: Workshop on Standardization for Soundscape Techniques II

Brigitte Schulte-Fortkamp, Cochair
Technical University Berlin, Inst. of Fluid Mechanics and Engineering Acoustics, Einsteinufer 25, D-10587 Berlin, Germany

Bennett M. Brooks, Cochair
Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, CT 06066

1:30

Working Groups, Part 3

The combination of physical acoustical measurements with scientific evaluation of perceptual responses to environmental sound,
known as soundscaping, is an essential method for the assessment and actualization of positive outdoor environments. Working groups
will be organized around the following areas of interest, as expressed by persons responding to the workshop announcement or at
meeting registration. Working groups will develop presentations on problems, solutions, and recommended actions. To further develop
and refine the methods of soundscaping the following are needed: (1) Catalog of correlations between physical parameters and

perceptual responses. (2) Standardization of a terminology lexicon of soundscape descriptors. (3) Standardization of measurement procedures. (4) Recommendations for perceptual evaluation and analysis. Working group discussions from the morning will continue, focusing on (1) what has been done in the past, (2) what should be done now, and (3) how may this be accomplished?

2:30

Final Group Presentation

The final plenary session of the workshop will include presentations by the moderators/recorders of all the working groups on their findings. Discussion will include proposed methods, means, and possible venues for follow-up and further action. The final set of recommendations will be published and distributed to all of the participants.

TUESDAY AFTERNOON, 5 JUNE 2007   GRAND BALLROOM A, 1:00 TO 5:00 P.M.

Session 2pPA

Physical Acoustics: Ultrasound in Condensed Matter, Neutrons, Nano-Materials, Magnetism

Albert Migliori, Cochair
Los Alamos National Lab., Los Alamos, NM 87545

Veerle M. Keppens, Cochair
Univ. of Tennessee, Materials Science and Engineering, Dougherty Hall, Knoxville, TN 37996

Chair's Introduction—1:00

Invited Papers

1:05

2pPA1. Bulk versus nanoscale WS$_2$: Finite size effects and solid state lubrication. J. L. Musfeldt, S. Brown (Dept. of Chemistry, Univ. of Tennessee, Knoxville, TN 37996, musfeldt@utk.edu), I. Mihut, J. B. Betts, A. Migliori (Los Alamos Natl. Lab. Los Alamos, NM), A. Zak, and R. Tenne (Weizmann Inst. of Sci., Israel)

Metal dichalcogenide nanostructures have recently attracted attention due to their unique closed-cage structures, hierarchy of length scales, and outstanding solid state lubrication behavior. To understand the finite size effects and tribological properties in these nanoscale materials, we measured the low temperature specific heat of layered and nanoparticle WS$_2$. Below 9 K, the specific heat of the nanoparticles deviates from that of the bulk. Further, the thermal response of the nested nanoparticles deviates from the usual T^3 dependence below 4 K because of both finite size effects and inter-particle-motion entropy. This separation of nanoscale effects from T^3 dependence can be modeled by assuming that the phonon density of states is flexible, changing with size and shape of the nanoparticle. We also invoke relationships between the low temperature T^3 phonon term, Young's modulus, and friction coefficient to assess the difference in the tribological properties. Based on this analysis, we conclude that the improved lubrication properties of the nanoparticles are extrinsic in origin.

1:30

2pPA2. Elastic constants, Blackman diagrams, and new lanthanide−actinide insights. Hassel Ledbetter (Mech. Eng. Dept., Univ. of Colorado, Boulder, Colorado 80309, Hassel.Ledbetter@colorado.edu)

Using monocrystal elastic constants and Blackman diagrams, one can infer material interconnections, physical-property trends, and knowledge about interatomic bonding. After reviewing briefly the f.c.c.-metal case, focus shifts to Ce, Pu, Th, with known monocrystal f.c.c. elastic constants, the Cij. For U and a few other lanthanide-actinide cases, one can deduce the f.c.c. Cij. From these results, for the lanthanides-actinides, there follow several conclusions: (1) elastic anisotropy ranges widely; (2) Poisson ratios are low; (3) interatomic bonding varies widely; (4) against other f.c.c. metals, all show unusual negative Cauchy pressures (three-body forces); (5) a strong covalent-bonding component occurs, perhaps related to localized (or semilocalized) f-electrons; (6) delta-Pu behaves most oddly; (7) delta-Pu may share some bonding features with f.c.c. alkali metals. [Work supported by DoE, NSF, State of Florida.]

1:55

2pPA3. Magnetoacoustic studies as a probe of electron systems. Alexey Suslov (NHMFL, 1800 E. Paul Dirac Dr., Tallahassee, FL 32310, souslov@magnet.fsu.edu)

The pulse-echo technique was used to probe the electron systems in solids. The measurements were performed in the frequency range 10–500 MHz, at temperatures down to 0.3 K and in the magnetic fields up to 18 T. (a) Acoustic quantum oscillations were investigated in URu_2Si_2, $AuZn$, and Sr_2RuO_4. Extracted parameters of the electron systems are in good agreement with data known from other experiments and theoretical calculations. Hydrostatic pressure effect on superconductive transition in Sr_2RuO_4 was studied. It is found that the sound speed increases, critical temperature and critical fields decreases, and transition becomes broader

under the pressure. Magnetic field dependencies of ultrasound attenuation are compared with theoretical predictions. (b) Surface acoustic waves were used for contactless measurements of conductivity in low dimensional structures. The value of complex ac conductivity was extracted from simultaneous measurements of the sound attenuation and velocity. Such measurements allowed to study, for example, mechanisms of conductivity in a dense array of SiGe quantum dots and localization of the 2D carries in GaAs/AlGaAs and Si/SiGe heterostructures in the extreme quantum limit. NHMFL is supported by the NSF Cooperative Agreement No. DMR-0084173 and the State of Florida. The ultrasonic research at the NHMFL is supported by the In-House Research Program.

2:20

2pPA4. Scattering methods applied to lattice dynamics in thermoelectric materials. Raphaël P. Hermann (Institut fur Festkörperforschung, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany)

Inelastic neutron scattering and x-ray scattering or nuclear resonance scattering are powerful techniques that yield insight into the microscopic origin of macroscopic material behavior. Progress in nuclear resonance scattering that allows element specific studies of lattice dynamics and, in particular, element specific measurements of the phonon density of states, atomic displacement parameters, and force constants, will be discussed. Recent results from lattice dynamics studies of thermoelectric materials, such as filled skutterudites, clathrates, and Zn_4Sb_3, will be presented in conjunction with results from macroscopic determinations of the thermo-dynamical properties of these materials, such as heat capacity or elastic constants obtained by ultrasound spectroscopy. [The European Synchrotron Radiation Facility is acknowledged for provision of the synchrotron radiation facility at beamlines ID18 and ID22N. Part of this research was carried out at the FRJ-II research reactor in Jülich, Germany.]

2:45

2pPA5. The charge ordering transition as probed by ultrasound. David Mandrus, Manuel Angst (Mater. Sci. and Technol. Div., Oak Ridge Natl. Lab., P.O. Box 2008, MS6056, Oak Ridge, TN 37831), Yanbing Luan, and Veerle Keppens (The Univ. of Tennessee, Knoxville, TN 37996)

Charge ordering (CO) phenomena are found in many correlated electron systems but in general are poorly understood. A classical example of CO is found in magnetite, Fe_3O_4, which undergoes a CO transition (the so-called Verwey transition) at 120 K. Magnetite is a poor model system for studying CO; however, despite nearly 70 years of study, there is still no generally accepted description of the transition. In recent years several new materials that display CO transitions have been discovered, including Fe_2OBO_3 and $LuFe_2O_4$. These new materials are proving to be far better model systems for studying CO than magnetite. In this talk the Landau theory of CO transitions will be reviewed, and the behavior of the elastic response of $LuFe_2O_4$ will be discussed in this context. [Work at ORNL supported by DOE BES Division of Materials Science and Engineering. Work at UT supported by NSF DMR-0506292.]

3:10−3:25   Break

3:25

2pPA6. Elasticity in metallic glasses. Takeshi Egami, Valentin Levashov (Dept. of Mater. Sci. and Eng. and Dept. of Phys., Univ. of Tennessee, Knoxville, TN 37996-1508), Rachel Aga, and Jamie Morris, J (Oak Ridge Natl. Lab., Oak Ridge, TN 37831)

The unique nature of elasticity in metallic glasses will be discussed with particular focus on the intrinsic anelasticity. Although the bulk modulus of a metallic glass is comparable in magnitude to that of the corresponding crystalline solid, the shear modulus is always lower by 20% to 30%. This is because shear deformation of a metallic glass produces not only the geometrical deformation but also local topological changes of atomic bonding. Even in the glassy state there is some local residue of liquidity due to local structural frustrations, and this makes the shear deformation in metallic glasses intrinsically anelastic. On this point glasses are fundamentally different from crystals. This intrinsic anelasticity results in anomalous dependence of shear modulus and anelastic loss on frequency and temperature, including the low-temperature two-level state behavior. Based upon computer simulation, neutron scattering experiment, and theory, the intrinsic nature of anelasticity will be shown to be a part of the big picture of the structure and dynamics of metallic glasses.

3:50

2pPA7. Pulse-echo ultrasound in pulsed magnetic fields. D. Rickel (Los Alamos Natl. Lab. and Natl. High Magnetic Field Lab., MS E536, Los Alamos, NM 87545, drickel@lanl.gov)

As the technology of high-speed data acquisition evolves along with the ability to record multi-megabyte continuous streams of data, the application of pulse-echo ultrasound to transient measurements have become possible. I will discuss a system that has been developed at the National High Magnetic Field Laboratory for measurements of sound speed and attenuation in materials subjected to intense pulsed magnetic fields. The magnetic field pulses have millisecond durations and can only be repeated a few times a day. We are able to process from a single magnetic pulse, a record of sound speed that reveals structural changes as a function of magnetic field. We are not alone in this effort; at least two other groups at other high magnetic field laboratories are using similar techniques [C. Proust *et al.*, Lab. National des Champs Magnetiques Pulses, Toulouse France, UMR 5147CNRS-INSA-UPS; Wolf *et al.*, "New Experimental Techniques for Pulsed Magnetic Fields-ESR and Ultrasonics," Physica B **294−295**, 612–17 (2001)]. This presentation is intended to acquaint the audience with the techniques of setting up a pulse-echo ultrasound system by taking advantage of modern digitizing systems available and computer based post processing using commercially available application software. I will present data from experiments that we have carried out and discuss the precision with which the sound speeds can be determined.

4:15

2pPA8. Thermal diffusivity of $Bi_2Sr_2CaCu_2O_8$ superconductor measured using open-cell photoacoustic technique: Effect of nano-Ag addition. Hasan A. Alwi, Mary A. B. Narreto, Kong Wei, and Roslan Abd-Shukor (School of Appl. Phys., Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia)

Thermal diffusivities of $Bi_2Sr_2CaCu_2O_8$ superconductor at room temperature with the addition of different amounts of nano Ag ($Bi_2Sr_2CaCu_2O_8$-Ag_x with $x=0$-1.0) were measured using the open-cell photoacoustic technique. The thermal diffusivity was obtained by analyzing the phase of photoacoustic signal of thermally thick samples instead of analyzing the amplitude signal of thermally thick-to-thin samples. Two methods of phase analysis of thermally thick samples were applied, that is, Calderon's method and the thermal diffusion model. The phase analysis is suitable where the dominanant mechanism contributing to the photoacoustic signal is thermal diffusion alone, which is true for $Bi_2Sr_2CaCu_2O_8$ superconductor. We have found that the thermal diffusivity increases with the amounts of nano-Ag. The values of thermal diffusivity obtained were between 0.012 50-0.001 28 and 0.016 06-0.002 30 cm^2/s for Ag_0 to $Ag_{1.0}$.

4:30

2pPA9. Determination of third-order elastic constants of langasite single crystals through force-frequency effect. Haifeng Zhang, Joseph A. Turner, Jiashi Yang (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, Lincoln, NE 68588-0526, hfzhang@bigred.unl.edu), and John A. Kosinski (U.S. Army RDECOM CERDEC, Fort Monmouth, NJ 07703-5211)

Langasite resonators are of recent interest for a variety of applications because of their good temperature behavior, good piezoelectric coupling, low acoustic loss, and high Q factor. Third-order elastic constants are responsible for many nonlinear effects associated with langasite resonators including the force-frequency effect. This effect describes the shift in reso-nant frequency a resonator experiences due to the application of a mechanical load. The determination of the third-order elastic constants through the force-frequency effect relies on an accurate theoretical understanding of this effect. In this article, expressions for the frequency shift of langasite resonators subjected to a pair of diametric forces are obtained using the perturbation integral in conjunction with the finite element method. Experimental measurements of the force-frequency effect for various langasite resonator samples with plano-plano configuration are also discussed. The third-order elastic constants of langasite single crystals are extracted based on the measured data in conjunction with the derived expressions. The results are compared with previous measurements of third-order elastic constants for langasite obtained using transit-time methods. [Work supported by ARO.]

4:45

2pPA10. Diffuse waves in solids: Can you hear the texture of a crystal? Andrew Norris (Mech. and Aerosp. Eng., Rutgers Univ., Piscataway, NJ 08854)

Diffuse waves are a steady state reverberation that partitions energy equally among modes of a system. In an isotropic solid of infinite extent the equipartition defines a well known equation for the spectral density d, which is proportional to the square of the frequency. The proportionality factor depends on the material properties through an average of the inverse third power of the wave speeds. This talk describes the analogous equation for diffuse wave density in a crystal. The first result is that the density d is a matrix or tensor quantity, which possesses the material symmetry of the solid. The properties of the matrix are discussed both for the forward problem of determining d for a given anisotropic material, and the inverse question of what information can be derived from d about the local material texture. The general form of the spectral density presented is related to the admittance matrix for a point force, as will be explained.

TUESDAY AFTERNOON, 5 JUNE 2007   GRAND BALLROOM B, 1:30 TO 4:30 P.M.

Session 2pPP

**Psychological and Physiological Acoustics: Pitch, Binaural and Spatial Hearing Suppression
(Poster Session)**

Christopher A. Brown, Chair

Arizona State Univ., Speech and Hearing Science, Box 870102, Tempe, AZ 85287-0102

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-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

2pPP1. Dominance region for pitch at low fundamental frequencies: Implications for pitch theories. Brian Moore, Brian Glasberg, Idriss Aberkane, Samantha Pinker (Dept. of Exp. Psych., Univ. of Cambridge, Downing St., Cambridge CB2 3EB, UK), and Candida Caldicot-Bull (Univ. of Cambridge, Cambridge CB3 9DP, UK)

The dominant region for pitch was measured for complex tones with low fundamental frequency (F0=35 and 50 Hz). The tones contained 59 harmonics, added in cosine or random phase. The harmonics were split into two groups; group A containing harmonics 1-K and group B containing harmonics $(K+1)$-N. On each trial, two successive complex tones were presented. In one, the harmonics in group A were shifted down by deltaF0 and the harmonics in group B were shifted up by deltaF0. In the other tone, the shifts were in the opposite direction. The tones were presented in random order and the subject had to indicate which had the higher pitch. The pitch judgements followed the components in group B for small K and the components in group A for large K. The frequency corresponding to the middle of the transition region was taken as the center of the dominant region, fdom. The value of fdom varied markedly

across subjects but typically corresponded to harmonic numbers above the 6th, for both phase conditions. Thus, the dominant harmonics were unresolved. These results indicate that resolution of individual harmonics is not the key factor determining the location of the dominant region. [Work supported by the MRC.]

2pPP2. Why pitch strength decreases with increasing harmonic number in complex tones. D. Timothy Ives and Roy D. Patterson (CNBH, Dept. Physiol., Development, and Neurosci., Univ. of Cambridge, Cambridge, UK)

A melodic pitch experiment was performed to demonstrate the importance of temporal resolution for pitch salience. The experiments show that notes with a low fundamental (75 Hz) and relatively few resolved harmonics support better performance than comparable notes with a higher fundamental (300 Hz) and more resolved harmonics. Two four-note melodies were presented to listeners and one note in the second melody was changed by one or two semitones. Listeners were required to identify the note that changed. There were four orthogonal stimulus dimensions: F0 (75 and 300 Hz), lowest frequency component (3, 7, 11, or 15), number of harmonics (2, 4, or 8), and degree of component rove (1 or 3). Performance decreased as the frequency of the lowest component increased for both F0s, but performance was better for the lower F0. The spectral and temporal information in the stimuli were compared using the Auditory-Image Model [Bleeck *et al.*, Acta Acust. **90**, 781–788 (2004)]. Peaks in the time-interval profile of the auditory image can explain the decrease in performance as F0, and spectral resolution increase. Spectral profiles of the same auditory images do not contain sufficient resolution to explain the performance. [Work supported by U.K. MRC (G0500221, G9900369).]

2pPP3. The effect of phase, fundamental, and lowest component on the perception of octave height. Ralph van Dinther and Roy D. Patterson (CNBH, Dept. Physiol., Development and Neurosci., Univ. of Cambridge, Downing St., Cambridge, CB2 3EG, UK)

Attenuation of the odd harmonics of a complex harmonic tone increases the perceived octave height and eventually leads to a shift of one complete octave in the perception. In a 2AFC experiment, we assessed how much the odd harmonics had to be attenuated to make a sound indistinguishable from its octave, first when the components were in cosine phase and then when the phase of the odd harmonics was shifted by $\frac{1}{3}\pi$ or $\frac{1}{2}\pi$. Thresholds were measured for four fundamental frequencies (31, 63, 125, 250 Hz) and four lowest harmonic numbers (2, 4, 8, 16). The results show that when the odd harmonics are phase shifted, the attenuation required to make the test sound indistinguishable from its octave decreases dramatically, from between 20 and 30 dB to less than 5 dB. A model based on the time-interval profile of the auditory image [Patterson, J. Acoust. Soc. Am. **98**, 1890–1894 (1995)] was used to explain the data; 95% of the predicted thresholds fall within one standard deviation of the empirical data. The spectral profile did not have sufficient resolution to predict the data and it could not explain the effect of the phase manipulation. [Work supported by the U.K. Medical Research Council (G0500221, G9900369).]

2pPP4. Temporal relationship between speech auditory brainstem responses and phonemes envelopes. Idrick Akhoun, Stephane Gallego, Evelyne Veuillet, and Lionel Collet (Laboratoire Neurosci. et Systemes Sensoriels, 50, avenue T. Garnier, F-69007 Lyon, France, idrick.akhoun@chu-lyon.fr)

Auditory brainstem responses evoked by synthetic phonemes (speech-ABR) were recorded on normal hearing subjects stimulated monaurally on the right ear at 60 dB sensation level. The speech-ABR consists of an onset response (OR) followed by the frequency following response (FFR), which results of the summation of phase locking underlying activities in the auditory cells of the brainstem nuclei. The FFR mimicked accurately the envelope of the stimulus, defined as the absolute value of the stimulus filtered at the cutoff frequency of the speech ABR spectrum. However, the FFR was more shifted relative to the vowel than the OR did relative to the onset of the stimulus, which is close to the cochlear delay usually reported. Little interindividual discrepancies were found. Delay values give information about the sites of generation of the speech-ABR. This method may help to objectivate the temporal encoding of phonemes in the brainstem. Hence the potential assessing of speech intelligibility, as temporal speech envelopes retain perfect intelligibility [R. Drullman, J. Acoust. Soc. Am. **98**, 1796–1798 (1995)]. [Work supported by CNRS.]

2pPP5. Experiments on perceiving the missing fundamental by using two harmonic components tone. Yutaka Iitomi and Takahide Matsuoka (Utsunomiya Univ., Yoto 7-1-2, Utsunomiya, Tochigi, Japan 321-8585, mt063203@cc.utsunomiya-u.ac.jp)

There are some methods on acoustic experiments for perceiving the missing fundamental. Listening to one of two harmonic components from each ear and listening to a complex tone of two harmonic components from each ear are the two examples. The latter method may perceive the combination tones also. Computational experiments of the above-mentioned two methods were done by using AIM (Auditory Image Model). The effect of the value of the fundamental and the effect of the number of harmonic components on extraction of the missing fundamental information were investigated. Finally, the result of the latter method was compared with the result of the former method and the effect of combination tones on extraction of the missing fundamental information was investigated. As a result of investigating many data, the fundamental could be perceived easier in the latter method (tone containing the combination tones) than in the former method. Its mechanism was considered.

2pPP6. Phase effects in F0 discrimination and temporal fine structure. Andrew J. Oxenham and Christophe Micheyl (Dept. of Psych., Univ. of Minnesota, N640b Elliott Hall, 75 East River Rd., Minneapolis, MN 55455-0344)

Pitch discrimination with only high-order (unresolved) harmonics is dependent on their relative phases and is thought to be based on the temporal envelope of the waveform after peripheral filtering. In contrast, when low-order harmonics are present, phase effects disappear, and pitch may be derived from the estimated frequencies of the individual resolved harmonics. The dichotomy between the envelope-based periodicity from unresolved harmonics and the temporal fine-structure (or spectrally) based estimates of individual frequencies from resolved harmonics forms the basis of the well-known two-mechanism theory of pitch. Recently, the early idea that listeners use temporal fine-structure cues in unresolved harmonics to estimate the overall F0 has been revived as a third intermediate mechanism to account for why accurate F0 discrimination, normally associated with resolved harmonics, can sometimes be achieved in conditions that produce phase effects. Here we show that it is not necessary to resort to a third mechanism. Instead, we show that the relative component phases affect the instantaneous frequency distributions in ways that could influence the coding of frequencies for individual, partially resolved, harmonics. Implications for the interpretation of temporal fine structure in pitch and other auditory functions, such as speech perception, are discussed. [Work supported by NIH R01DC05216.]

2pPP7. Neural correlates of auditory streaming with unresolved harmonics. Ifat Yasin (Dept. of Exp. Psych., South Parks Rd., Univ. of Oxford, Oxford OX1 3UD, UK) and Christopher Plack (Lancaster Univ., Lancaster LA1 4YF, UK)

The magnetoencephalographic negative wave N1m may be used as an indicator of auditory stream segregation [Gutschalk *et al.*, J. Neurosci. **25**, 5382–5388 (2005)]. When a repeating sequence of a pure-tone triplet (ABA-) is perceived as a single stream, the N1m amplitude is relatively

small and consistent with the inter-stimulus-interval (ISI) between successive tones. When the sequence is perceived as segregated, the N1m amplitude is relatively large and consistent with the longer ISI between tones of each segregated stream. The present study used the electroencephalographic equivalent, wave N1, to investigate stream segregation based solely on temporal cues. Each triplet was comprised of complex tones with an f0 = 62.5, with unresolved harmonics summed either in sine phase, or in alternating sine-cosine phase, with a pitch corresponding to f0 and 2* f0, respectively. The effect of disrupting segregation was investigated by introducing increasing amounts of temporal jitter imposed by reducing the ISI for a complex within a triplet. Initial analyses tend to suggest that N1 amplitude is increased if the conditions for segregation are appropriate (when presentation rate is increased). Furthermore, N1 amplitude decreases as the amount of jitter increases, consistent with a disruption of segregation. [Work supported by ESRC and MRC.]

2pPP8. The effect of fundamental frequency in simulated electric-acoustic hearing. Christopher A. Brown and Sid P. Bacon (Dept. of Speech and Hearing Sci., Arizona State Univ., P.O. Box 870102, Tempe, AZ 85287-0102)

Individuals with residual hearing restricted to low frequencies are candidates for electric-acoustic stimulation (EAS). When low-frequency acoustic information is added to either real or simulated high-frequency electric stimulation, speech recognition often improves dramatically. This may reflect the availability of fundamental frequency (f0) information in the acoustic region. We tested this directly. In particular, we examined the relative contributions of f0 and the amplitude envelope of speech by replacing the low-frequency speech with a tone that was modulated either in frequency to track the f0 of the speech, in amplitude with the extracted envelope of the low-frequency speech, or both. A female talker was combined with various backgrounds and processed with a four-channel vocoder to simulate electric hearing. Across all backgrounds, intelligibility improved significantly when a tone tracking f0 was added to vocoder stimulation and further still when both f0 and amplitude envelope cues were applied. These results confirm the importance of f0 information (at least under simulated EAS) and indicate that significant information can be provided by a tone that tracks f0 or f0 and amplitude envelope cues. [Work supported in part by NIH.]

2pPP9. The effect of interaural differences of time on speech intelligibility in simulated electric-acoustic hearing. Christopher A. Brown and Sid P. Bacon (Dept. of Speech and Hearing Sci., Arizona State Univ., P.O. Box 870102, Tempe, AZ 85287-0102)

Individuals with residual hearing restricted to frequencies below about 500–750 Hz have recently been implanted with a relatively short electrode array designed to preserve as much of the residual hearing as possible. Those who also have some low-frequency hearing in the non-implanted ear have the potential to benefit from interaural time differences (ITD). The goal of the current experiments was to determine whether such benefit exists in simulations of implant listening and how it is influenced by sensation level (SL) in the low frequencies. Presentation was via headphones. Target and distractor speech were low-pass filtered at 500 Hz. The target was presented diotically and the distractor was presented with a 0-s (spatially centered image) or 600-s (image to the right) ITD. The target and distractor were also processed by a four-channel vocoder and presented to the left ear only. The SL of the low-frequency speech was varied both symmetrically and asymmetrically. In the symmetric conditions (15–60 dB SL), intelligibility improved by as much as 40% when the target and distractor were spatially separated. In the asymmetric conditions (60 dB in one ear, 20–50 dB in the other), approximately 15%–20% improvement was observed in all conditions. [Work supported by NIH and NOHR.]

2pPP10. Laboratory rats do not use binaural time cues to localize sound. Christina M. Wesolek,[a] , Gimseong Koay, and Henry E. Heffner (Dept. of Psych. MS 948, Univ. of Toledo, 2801 W. Bancroft St., Toledo, OH 43606, Henry.Heffner@utoledo.edu)

The use of binaural time and intensity cues to localize sound can be investigated by determining the ability of a subject to localize pure tones in a free-field. Specifically, the ability to localize low-frequency tones, which do not produce an intensity difference at the two ears, demonstrates the use of the binaural phase (time) cue whereas the ability to localize high-frequency tones, to which the auditory system cannot phase lock, demonstrates the use of the binaural intensity-difference cue. Because previous studies of the laboratory rat (*Rattus norvegicus*) disagreed on the highest frequency that could be localized using binaural phase difference, the ability of rats to localize pure tones was reexamined. The results indicated that, contrary to previous studies, laboratory rats are not able to localize low-frequency tones even when they are amplitude modulated. Thus, it appears that laboratory rats are unable to use the binaural time-difference cue to localize sound. Because the previous studies were conducted before the widespread availability of spectrum analyzers, it is possible that their results were due to the presence of high-frequency harmonics in their low-frequency tones. These results have relevance for the anatomical and physiological study of binaural processing in laboratory rats. [a] Presently at Disney's Animal Kingdom, Lake Buena Vista, FL.

2pPP11. Auditory evoked potential hearing measures of a group of bottlenosed dolphins (*Tursiops truncatus*). Kristen A. Taylor, Paul E. Nachtigall (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734), Mats Amundin, Pernilla Mosesson, Sunna Edberg, and Stina Karlsson (Kolmrdens Djurpark AB, SE-61892, Kolmrden, Sweden)

The hearing of the bottlenosed dolphins at Kolmrdens Djurpark, Sweden, was examined using a portable auditory evoked potential (AEP) system in order to examine the hearing capabilities of dolphins housed in quiet environments with an emphasis on age-related hearing variation. Audiograms were measured on two dolphins and modulation rate transfer functions (MRTF) were also measured on two animals. A complete audiogram (4–200 kHz) was obtained for the oldest dolphin, a 34 year old female named Vicky. Her audiogram showed no evidence of high-frequency hearing loss. Notwithstanding this finding, her click evoked auditory brainstem response (ABR) revealed a marked increase in interpeak latency that could be indicative of an underlying hearing defect. The audiogram obtained from for the male dolphin, a 24 year old named Pichi (measured from 4–107 kHz), demonstrated profound hearing loss at all frequencies. The measured MRTF results indicated typical odontocete following responses to rates exceeding 1000 Hz.

2pPP12. Hearing screening in bottlenose dolphins using single and multiple auditory evoked potentials. James J. Finneran (U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Ctr., Code 2351, 49620 Beluga Rd., San Diego, CA 92152), Dorian S. Houser (Biomimetica, Santee, CA 92071), Dave Blasko, Christie Hicks, Jim Hudson, Mike Osborn (The Mirage Dolphin Habitat, Las Vegas, NV 89109), and Kevin M. Walsh (Gulf World Marine Park, Panama City Beach, FL 32413)

Auditory thresholds were estimated in four bottlenose dolphins (*Tursiops truncatus*) by measuring auditory evoked responses to single and multiple sinusoidal amplitude modulated tones. Subjects consisted of two males and two females with ages from 4 to 22 years. Testing was conducted in air using a jawphone transducer to couple sound into the subjects lower right jaw. Carrier frequencies ranged from 10 to 160 kHz in one-half octave steps. Amplitude-modulated stimuli were presented individually and as the sum of four, five, and nine tones with unique carrier and modulation frequencies. Evoked potentials were noninvasively recorded using surface electrodes embedded in silicon suction cups. The presence or absence of an evoked response at each modulation frequency was assessed

by calculating the magnitude-squared coherence from the frequency spectra of the recorded epochs. All subjects exhibited traditional U-shaped audiograms with upper cutoff frequencies above 113 kHz. The time required for threshold estimates ranged from 30–45 min for single stimuli and 4–8 min for nine simultaneous stimuli. Agreement between thresholds estimated from single stimuli and multiple, simultaneous stimuli was generally good, suggesting that multiple stimuli may be used for quick hearing assessment when time is limited. [Work supported by ONR.]

2pPP13. Effects of ear coverage and reflected sound on the localization of sound. Angelique Scharine, Tim Mermagen, Justin MacDonald, and Mary Binseel (Army Res. Lab., Human Res. and Eng. Directorate AMSRD-ARL-HR-SD; APG, MD 21005-5425)

Previous research has shown that listeners wearing helmets with greater ear coverage perform worse on localization tasks. However, the helmets studied were selected from those commonly in use and differed in suspension systems and profiles. Three versions of the same helmet design, differing only in the ear coverage (0%, 50%, 100%), were compared in this study. Sounds were presented from 12 azimuthal locations spaced 30 deg apart in the presence of a white noise masker. Twelve listeners completed a localization task while wearing each of the helmets as well as with no helmet in two different acoustic environments. As ear coverage increased, localization performance decreased. The effects of early reflections, though small, were consistent across the different coverage conditions. The effect of ear coverage and its interaction with the acoustic environment will be discussed in terms of its significance for real-world environments. In addition, head-related transfer-functions (HRTFs) were measured for each listener in each of the head conditions. A model of localization performance based on changes to the HRTF will be described and applied to the collected data.

2pPP14. Multiple echoes precedence effect. Gongqiang Yu and Ruth Y. Litovsky (Waisman Ctr., Univ. of Wisconsin—Madison, Madison, WI 53705)

The precedence effect has been explored extensively using a lead and a single lag (simulated echo). For click stimuli, when the lead-lag delay is a few ms, discrimination of directional changes in the lag are poorer than for the lead. This study was aimed at determining whether the effects would be upheld in a three-source paradigm and to explore the extent to which the three sources interact. Three binaural click pairs (Lead, Lag1, and Lag2) were presented over headphones. Discrimination of directional changes in each click pair was measured using an ABX, 2AFC task, for delays of 0–130 ms. Results replicated classic findings with a single lag, but also showed that (1) for delays above 16 ms a temporal order effect resulted in poorer lead discrimination and better Lag2 discrimination; (2) echo thresholds were complicated by the addition of Lag2; and (3) an interaction was observed between Lag1 and Lag2 at very brief time delays such that summing localization resulted in improved discrimination of lagging sound sources. These results suggest that in a multi-echo situation the directional cues from echoes that occur after the classic lead-lag pair can contribute in the localization process and introduce complexities.

2pPP15. Spatial generalization in short-term perceptual adaptation to novel auditory spaces. I-Fan Lin, Timothy Streeter, Barbara Shinn-Cunningham (Auditory Neurosci. Lab., Dept. of Cognit. and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215, ifan1976@cns.bu.edu), Nat Durlach, and Lorraine Delhorne (Boston Univ., Boston, MA 02215)

Previous work shows that humans can rapidly adapt to novel mappings between auditory spatial cues and exocentric location and localize accurately with the novel mapping. Here, experiments using (1) a rotation and

(2) a linear magnification of auditory space examined generalization of spatial adaptation to untrained or neglected regions of space. Individual head-related transfer functions (HRTFs) were used to simulate different source locations using headphones. Each experimental session was broken into three blocks: initial normal cues (measuring baseline performance), altered cues (showing adaptation), and final normal cues (measuring an adaptation aftereffect). In the altered cues trials, a source at X degrees was simulated using HRTFs that normally correspond to a source at (1) X +30 deg (rotation) or (2) 2X deg (linear magnification). After obtaining subjects' localization responses to the sound stimulus, feedback was given using a light-emitting diode located at the correct position (X deg). Results of these experiments show the extent to which training subjects to adjust their localization responses in one region of space affects how they localize sources in adjacent regions, and provide a measure of the spatial kernel that drives spatial auditory adaptation.

2pPP16. Interactive acoustical intimacy exploration. Bobby Gibbs II, Jonas Braasch, and Ted Krueger (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

A novel methodology to explore salient cues for acoustical intimacy is presented. Subjects are instructed to modify a set of sliders in a virtual acoustical environment to correlate with three intimacy settings: low, medium, and high. As the experiment proceeds, subjects are allowed to audition their choices and make modifications to achieve what they feel to be optimal settings. At the conclusion of the interactive phase, subjects are instructed to rank the (as yet unknown to them) parameters in terms of their overall effect on intimacy. Unknown to the subjects, each slider correlates to an acoustical parameter: direct-to-reverberant energy ratio, high-frequency attenuation, room size, and volume. Results from the experiment are discussed relative to concert-hall acoustics and virtual spaces.

2pPP17. Visual and auditory hemispheric cuing in horizontal sound localization. Norbert Kopčo (Dept. of Cognit. and Neural Systems, Boston Univ.; Percept. and Cognition Lab, Tech. Univ. of Košice, Slovakia, kopco@bu.edu), Beáta Tomoriová, and Rudolf Andoga (Tech. Univ. of Košice, Slovakia)

Experiments were performed to determine 1) whether *a priori* information influences sound localization, and 2) whether the influence depends on the modality through which the information is provided. Localization performance was measured for transient auditory stimuli originating in the frontal horizontal plane. In most runs, a cue preceded the stimulus and indicated (correctly or incorrectly) the hemisphere (left vs. right) from which the subsequent target arrived. The cues differed by modality and the cue-to-target onset asynchrony (SOA: 400 to 1600 ms). The listeners were instructed to focus their attention to the cued side. A follow-up experiment evaluated the effect of eye fixation on the visual-cue performance. Results show modality-dependent effects of cuing in terms of both bias and standard deviation in responses. A cue that indicated an incorrect side biased the responses towards its side in both modalities, while a visual correct-side cue caused either no bias or a lateral bias. A visual cue that indicated an incorrect side also caused an increase in the standard deviations at the largest SOA, suggesting that the mechanisms that control auditory spatial attention are modality dependent and that they operate on the time scale of seconds. [Work supported by the Slovak Science Grant Agency.]

2pPP18. Characteristics of visually-induced auditory spatial adaptation. I-Fan Lin (Dept. of Cognit. and Neural Sys., Boston Univ., 677 Beacon St., Boston, MA 02215), Norbert Kopčo (Boston Univ., Boston, MA 02215), Jennifer M. Groh (Duke Univ., Durham, NC), and Barbara G. Shinn-Cunningham (Boston Univ., Boston, MA 02215)

In humans, visual and auditory spatial information arise in different reference frames, with auditory information being head-centered based on interaural cues, whereas visual information derives from the retinal activation, an eye-centered signal. This study investigated whether visually-guided recalibration of auditory space occurs in a head-centered, eye-centered or a hybrid (combination of both) reference frame. Subjects made saccades to combined auditory-visual targets. The visual component of the target was displaced laterally by 5°. Interleaved auditory-only trials served to evaluate the effect of the mismatched visual stimulus on auditory space. To dissociate head- from eye-centered reference frames, the initial fixation position of the eyes was varied. The results reflected contributions of both coordinate frames. First, the magnitude of the induced shift was affected by the eye-centered location of the targets, but occurred at a consistent location in head-centered space. Second, when subjects were trained using one fixation position and tested using a different fixation position, the magnitude of the induced shift was reduced, but the target locations that showed the greatest shift did not move. The hybrid nature of these effects is reminiscent of recent demonstrations of eye position modulation in the auditory pathway of non-human primates. [Work supported by NIH]

2pPP19. The role of suppression in temporal masking effects. Lata A. Krishnan and Elizabeth A. Strickland (Dept. of Speech, Lang. and Hearing Sci., Purdue Univ., 1353 Heavilon Hall, 500 Oval Dr. West, Lafayette, IN 47907)

The temporal effect refers to the change in signal-to-masker ratio at threshold for a short-duration tone as it is delayed from the onset of a longer-duration masker. Evidence suggests that this phenomenon is related to changes in gain in the cochlea due to acoustic stimulation by the masker. Research has suggested that for off-frequency maskers, the temporal effect is consistent with a decrease in suppression, which may depend on a decrease in gain at the suppressor frequency. Most previous studies have utilized simultaneous masking paradigms to measure the temporal effect. The present study used tonal maskers in a forward masking paradigm, and measured temporal masking curves to elucidate the role of suppression in temporal masking effects. A previous study had also showed that the pattern of the temporal effect with level depended not only on the degree of hearing loss at the signal frequency, but also on the degree of hearing loss above the signal frequency. The present study examined this further by using listeners with normal hearing as well as listeners with a notched cochlear hearing loss. Results will be analyzed in terms of changes in gain at signal and suppressor frequencies. [Work supported by Lions Club McKinney Outreach Award.]

TUESDAY AFTERNOON, 5 JUNE 2007　　　　　　　　　　　　　TOPAZ ROOM, 1:30 TO 5:15 P.M.

Session 2pSA

Structural Acoustics and Vibration: Launch Vehicle and Space Vehicle Acoustical and Vibration Environments and General Topics in Structural Acoustics and Vibration

James E. Phillips, Chair
Wilson, Ihrig and Associates, Inc., 5776 Broadway, Oakland, CA 94618-1531

Chair's Introduction—1:30

Invited Papers

1:35

2pSA1. Derivation of aeroacoustic environments for the crew exploration vehicle. Michael Yang (ATA Eng., Inc., 11995 El Camino Real, Ste. 200, San Diego, CA 92130)

The Crew Exploration Vehicle (CEV) is designed to provide safer, more reliable space transportation capability in the quest for human space exploration. This article covers the development of an external aeronoise model which was fit to generic and Saturn/Apollo configuration wind tunnel data, and anchored to Atlas and Titan flight data. The predictions for a combined aerodynamic/propulsion system load case while firing the Launch Abort System (LAS) engines during flight is also presented, along with SEA methods for predicting internal responses.

1:55

2pSA2. Prediction of liftoff acoustic environments for launch pads with covered ducts. Michael Yang (ATA Eng., Inc., 11995 El Camino Real, Ste. 200, San Diego, CA 92130)

The acoustic environment experienced by a launch vehicle is typically most severe during the liftoff phase. The launch pad can be designed with a deflector to redirect the exhaust plume into a covered duct. This is done to reduce the fluctuating pressure field experienced by the vehicle. The Eldred method is frequently used to predict liftoff environments, but does not specify how to model the effects of the covered duct. Techniques used to model the covered duct are reviewed and compared in this paper.

2:15

2pSA3. High-frequency radiation from edges and drive points on vibrating panels. Donald Bliss and Linda Franzoni (Dept. of Mech. Eng., Duke Univ., Durham, NC 27708)

High-frequency radiation from subsonic waves on vibrating panels is shown to have directivity characteristics that can be expressed analytically by a limited set of parameters. Far-field mean-square pressures are independent of the panel length, indicating the radiation is associated with the edges and drive points. A Fourier transform approach to surface velocity shows that it can be reinterpreted in physical space in terms of singularity functions at the edges, namely the delta function and its derivatives. This interpretation leads to monopole, dipole, quadrupole, etc., edge radiators with relative strengths that depend on surface-wave Mach number. The radiation cannot be explained simply in terms of uncancelled volumetric sources at edges and drive points; this often-stated idea is not correct except under very restricted circumstances. A correct physical interpretation of the radiation is provided both in physical space and transform space. Basic directivity patterns are identified, associated with right and left traveling waves and the correlation between them at edges and drive points. The role of boundary conditions at the panel edge is illustrated, as are the effects of forcing. The analytical characterization of the radiation is shown to be particularly straightforward in the high-frequency broadband limit.

2:30

2pSA4. Control of aircraft interior noise using heterogeneous (HG) blankets. Kamal Idrisi, Marty Johnson, and James Carneal (Vib. and Acoust. Labs., Virginia Tech., 136 Durham Hall, Blacksburg, VA 24061, idrisi@vt.edu)

This study is concerned with the passive control of vibration and sound radiation of interior cabin noise in aircraft at low frequencies (<500 Hz) using heterogeneous (HG) blankets. HG blankets consist of poroelastic media with small embedded masses, which act similar to a distributed mass-spring-damper system. HG blankets have shown significant potential to reduce low-frequency radiated sound from structures, where traditional poroelastic materials have little effect. A mathematical model of a double panel system with an acoustic cavity and HG blanket was developed using mobility and impedance matrix methods. Theoretical predictions are validated with experimental measurements. Results indicate that proper tuning of the HG blankets can achieve in broadband reductions in sound transmission through the double panel system with less than 10% added mass. Future work includes expanding the model and experiment to multiple panel systems.

2:45

2pSA5. Steady-state vibration absorption using undamped linear substructures. Ilker Murat Koc (Dept. of Mech. Eng., Istanbul Tech. Univ., Inonu Cad. No. 87 Gumussuyu, Istanbul, Turkey 34437, ilker.koc@itu.edu.tr), Adnan Akay (Carnegie Mellon Univ., Pittsburgh, PA 15213), and Antonio Carcaterra (Univ. of Rome, 00184, Rome, Italy)

This study extends the previous applications of linear energy sinks to steady-state vibration absorption. First developed for the absorption of transient vibrations, linear energy sinks also act as effective vibration absorbers under steady-state excitation. However, their effectiveness depends on the value of excitation frequency with respect to the natural frequency of the primary structure around which the frequencies of the attached oscillators of energy sink are distributed. Vibratory energy is absorbed efficiently for excitation frequencies above the natural frequency of the primary structure, while absorption is not as effective at lower excitation frequencies. [Research carried out while AA served at NSF.]

3:00

2pSA6. Diffraction of acoustic waves and embedding theory. Richard V. Craster, Elizabeth A. Skelton (Dept. of Mathematics, Imperial College, London, UK), and Andrey V. Shanin (Moscow State Univ., Russia)

The diffraction and scattering of sound waves by sharp edges and corners is a ubiquitous feature of acoustic interactions with structures. This talk is on a technique called embedding which is a relatively new, and under used, idea in diffraction theory. The fundamental idea of embedding is that one only ever solves a single master, canonical, scattering problem (or set of problems). Thereafter, to extract the far-field behavior for any plane wave incidence on the same geometry, one only manipulates results from this master problem. The result is that quantities (directivities) previously dependent upon two parameters (incident and observed angle) become factorized into products of a function of a single variable, together with a simple trigonometric term; this facilitates rapid numerical evaluation. In principle, this should revolutionize many scattering calculations in every area where diffraction occurs as one need only evaluate the directivities from the set of canonical problems once. Given these directivities one manipulates them to generate solutions for more general problems, rather than continually recalculating and re-evaluating. At present, the theory has some limitations and the talk will cover these and describe how they can be overcome.

3:15–3:30   Break

3:30

2pSA7. The energy flow for guided waves at and inside the surface of an ensonified fluid loaded elastic cylindrical shell via the Poynting vector field: Preliminary results. Cleon E. Dean (Phys. Dept., P.O. Box 8031, Georgia Southern Univ., Statesboro, GA 30460-8031, cdean@GeorgiaSouthern.edu) and James P. Braselton (Georgia Southern Univ., Statesboro, GA 30460-8093)

The Poynting vector field is used to show preliminary results for the energy flow both at the surface and inside an ensonified fluid loaded elastic cylindrical shell for both forward and retrograde propagating waves. The present work uses a method adapted from a simpler technique due to Kaduchak and Marston [G. Kaduchak and P. L. Marston, ''Traveling-wave decomposition of surface displacements associated with scattering by a cylindrical shell: Numerical evaluation displaying guided forward and backward wave properties,'' J. Acoust. Soc. Am. **98**, 3501–3507 (1995)] to isolate unidirectional energy flows.

3:45

2pSA8. Low-wave-number turbulent boundary layer wall pressure measurements from vibration data on a cylinder in pipe flow. William K. Bonness and Dean E. Capone (Appl. Res. Lab., The Penn State Univ., State College, PA 16804)

Low-order vibration modes of a cylinder subjected to a fully developed turbulent boundary layer (TBL) in a pipe are measured and used to determine the low-wave-number content of energy in the boundary layer. The experiment is conducted using a 6-in.-diameter, thin-walled, aluminum test section filled with water flowing at 20 ft/s. Coupling of certain cylinder modes in the axial and radial directions allows use of the inverse method to determine the wave-number levels of TBL wall normal pressure and wall shear stress in the flow direction. These measurements are used to evaluate commonly used empirical models of the TBL wave-number pressure spectrum attributed to Corcos (1963) and Chase (1987) at lower wave numbers than have previously been reported. Preliminary results from several higher order cylinder modes confirm prior experimental findings that the Corcos model overestimates the low-wave-number pressures for these conditions by nearly 20 dB. The Chase model provides estimates within 3 dB of the measured low-wave-number pressure spectrum. [Work supported by Navsea 073R.]

4:00

2pSA9. Rayleigh waves guided by topography. Samuel D. M. Adams, Richard V. Craster, and Duncan P. Williams (Dept. of Mathematics, Imperial College London, South Kensington Campus, London, SW7 2AZ UK)

It would not be far-fetched to say that the work of Lord Rayleigh on surface guided waves has had fundamental and far-reaching effects upon modern life and many things we take for granted today, stretching from mobile phones through to the study of earthquakes. Many of these things take advantage of surface waves that are topographically guided thereby allowing energy to be carried in specific directions along some topography of the surface. Much of the emphasis has so far been placed on devices which are essentially thin rectangular ridge-like defects upon the surface. Indeed, experiments with these so-called SAW devices have demonstrated the presence of trapped waves. It remains to provide any convincing theory that captures this trapping effect for surface waves—this is our aim. We seek to resolve the issue of whether topography can support a trapped wave, whose energy is localized to within some vicinity of the topography, and to explain physically how trapping occurs. The trapping is first addressed by developing an asymptotic scheme that exploits a small parameter associated with the surface topography. We then provide numerical evidence to support results obtained from the asymptotic scheme; however, no rigorous proof of existence is presented.

4:15

2pSA10. Limiting performance and optimal synthesis of elastic-inertial multi-degree-of-freedom vibration isolation systems. Vyacheslav Ryaboy (Newport Corp., 1791 Deere Ave., Irvine, CA 92606)

This paper is concerned with optimal problems that seek to find the best vibration isolation performance achievable by a linear mechanical system with arbitrary structure and number of degrees-of-freedom. The solution is presented in form of inequality relating isolation efficiency, frequency domain, total mass, and overall static or quasi-static stiffness. The inequality is valid for all mechanical systems of the class, however complicated these systems may be. Partial optimal problems, such as mass minimization of a system with prescribed transmissibility function, are solved as intermediate steps in deducing the limiting performance inequalities. Synthesis of a vibration isolation system with prescribed properties is a closely related problem. It has multiple solutions that are presented in a systematized way; the class of the simplest systems allowing for a unique solution is defined.

4:30

2pSA11. A comparison between one-sided and two-sided Arnoldi-based model order reduction (MORe) techniques for fully coupled structural-acoustic analysis. R. Srinivasan Puri, Denise Morrey (School of Technol., Oxford Brookes Univ., Wheatley, Oxford, OX33 1HX, UK), and Jeffrey L. Cipolla (ABAQUS Inc., Providence, RI 02909-2499)

Reduced order models are developed for fully coupled structural-acoustic unsymmetric matrix models, resulting from Cragg's displacement/pressure formulation, using Krylov subspace techniques. The reduced order model is obtained by applying a *Galerkin* and *Petrov-Galerkin* projection of the coupled system matrices, from a higher dimensional subspace to a lower dimensional subspace, while matching the mo-

ments of the coupled higher dimensional system. Two such techniques, based on the Arnoldi algorithm, focusing on one-sided and two-sided moment matching, are presented. To validate the numerical techniques, an ABAQUS coupled structural-acoustic Benchmark problem is chosen and solved using the direct approach. First, the physical problem is modeled using ANSYS FE package and compared with closed form solutions. Next, ANSYS results are compared with nodal velocities obtained by generating reduced order models via moment matching. The results show that the reduced order models give a very significant reduction in computational time, while preserving the desired accuracy of the solution. It is also shown that the accuracy of the one-sided method could be further improved by using two-sided methods, where the Arnoldi vectors are *optimized* for chosen outputs. The accuracy of the approaches, convergence models, and computational times are compared.

4:45

2pSA12. Unsteady pressures on a ship hull. Michael Goody, Theodore Farabee, and Yu-Tai Lee (Naval Surface Warfare Ctr., Carderock Div., West Bethesda, MD 20817-5700)

Unsteady surface pressure measurements at locations distributed on the surface of a ship model hull and flow visualization of the bow wave and free surface are discussed. The pressure measurement locations are distributed over the hull surface from 15% to 70% of the model length. There are several additional pressure measurement locations on the hull surface adjacent to the bow wave. The measurements were performed in the David Taylor Model Basin using a ship model that is approximately 40 ft long and representative of a naval combatant at Froude numbers from 0.14 to 0.44. The measured surface pressure spectra are compared to predictions done using a RANS-statistical method and predictions done using an empirical formulation that is based on historical, flat-plate, surface pressure spectra. Scaling of the surface pressure spectrum and limited spatial coherence measurements are discussed. [Work sponsored by Office of Naval Research, Code 334.]

5:00

2pSA13. Generation and the measures of impact sound by stick-slip phenomenon at beam joint in buildings (the first report). Yasuaki Hayashi (Tasca Techno Corp., 2-10-4 Furou-Cho, Naka-Ku, Yokohama, 231-0032 Japan) and Hidemaro Shimoda (Koto-Ku, Tokyo, 135-8530 Japan)

It is often happened to be annoyed by the sound of an uncertain cause in the completed buildings. The phenomenon might be repeated indefinitely though most of these sounds are canceled with the passage of time. This paper reports a series of processes from the investigation of the cause, to effective measures on the case in a high rise housing complex. And the ascertained cause is thought to give a systematic explanation to this kind of uncertain sound in buildings. The outline of this paper is as follows. (1) It paid attention to the vibration wave form (phase) to which it propagated in the building frame to specify the position in the plane of the slab, the vibration accelerometers were arranged in two dimensional distribution on the slab, and the source was ascertained adequately. (2) It was clearly specified that the cause of generation was "stick-slip phenomenon" in the beam joint part from the wave form of the generated sound and its spectrum of the vibration. (3) Appropriate measures construction of completely controlling movement in this case was executed by being able to specify the cause of generation and the effect was confirmed.

Session 2pSC

Speech Communication: Frontiers of Spectrum Analysis with Speech Applications

Sean A. Fulop, Chair

California State Univ., Fresno, Dept. of Linguistics, 5245 N. Backer Ave., Fresno, CA 93740-8001

Chair's Introduction—1:30

Invited Papers

1:35

2pSC1. Complex peak time-frequency representations. Douglas Nelson   (Dept. of Defense, Fort Meade, MD 20705)

Complex representations encode twice as much information as representations encoding energy alone. Complex surfaces may be manipulated by conventional complex arithmetic operations to extract information about individual signal components. Presented are two representations. The first is an energy distribution representing energy and instantaneous frequency as the surface magnitude and phase, and the second is a signal distribution in which the instantaneous value of the signal at each time is distributed linearly in frequency. In each representation, the surface is first processed to concentrate the surface along curves representing the instantaneous frequencies of individual signal components, the surface at each time is then reduced to a short vector representing only the complex surface peaks with the largest magnitude. For the energy distribution model, the peak representation may be used to detect, track, and process individual signal components. For the signal distribution model, it is demonstrated that noise and interfering signals may be removed and an estimate of the clean signal may be recovered by a simple summation process. Moreover, the recovered clean signal is free of tonal artifacts produced by methods, such as spectral subtraction.

2:00

2pSC2. Efficient spectral measures for automatic speech recognition. Douglas D. O'Shaughnessy   (INRS-EMT, 800 de la Gauchetiere West, Ste. 6900, Montreal QC, Canada H5A 1K6)

It is well known that automatic speech recognition (ASR) requires good spectral analysis in order to have successful ASR accuracy. A wideband spectrogram seems to contain all the needed acoustic information to map any given speech signal into its corresponding sequence of phonemes. (For ASR, language models are often used to augment acoustics, but here we will limit ourselves to acoustic analysis.) Various methods beyond the basic Fourier transform have found success in ASR, e.g., linear predictive analysis, wavelets, and mel-frequency cepstra (MFCC). These have all been focussed on extracting an efficient set of spectral parameters to facilitate phonetic discrimination. Part of the difficulty is separating spectral envelope information from excitation parameters, as variations in pitch are largely viewed as orthogonal to phoneme recognition. Another complicating factor is that amplitude and frequency scales in speech production and perception are better modeled as nonlinear (unlike the linear, fixed-bandwidth approach of Fourier transforms). Modern ASR techniques are far from optimal, as the front-end data compression yielding MFCCs, for a basic 80-ms phoneme, typically has more than 100 parameters, to distinguish among approximately 32 phonemes (a 5-bit choice). We will investigate various ways to render ASR analysis more efficient. [Work supported by NSERC-Canada.]

2:25

2pSC3. Representing sound energy, phase, and interference using three-dimensional signals. Pantelis N. Vassilakis   (De Paul Univ., 2350 N. Kenmore Ave., J. T. Richardson Library 207, Chicago, IL 60614, pantelis@acousticslab.org)

As graphic representations of vibrations/waves, sound signals capture only selected attributes of the phenomenon they represent. Assuming equivalence between signals and sound waves obscures the fact that 2-D signals are unfit to a) represent wave-energy quantities consistently across frequencies, b) account for the phase flip and alternating positive/negative amplitude values in modulated waves with AM-depth >100%, and c) represent the energy content of interference. The proposed sound-signal representation is based on the complex equation of motion describing a wave. It results in spiral sine signals and twisted-spiral complex signals, similar to complex analytic signals, with the imaginary component of the complex equation of motion representing the signal envelope's argument (phase). Spiral sine signals offer a consistent measure of sine-wave energy across frequencies, while twisted-spiral complex signals account for the negative amplitudes observed in modulated signals, mapping the modulation parameters onto the twisting parameters. In terms of interference, 3-D signals illustrate that amplitude fluctuations and the signal envelopes that describe them are not just boundary curves but waves that trace changes in the total instantaneous energy of a signal over time, representing the oscillation between potential and kinetic energies within a wave. Examples of 3-D animations illustrating the proposed signals are presented.

3098   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3098**

2:50–3:00   Break

3:00

2pSC4. Inverting reassignment. Marcelo O. Magnasco (Rockefeller Univ., 1230 York Ave., New York NY 10021)

The short time Fourier transform has an extremely useful feature: its integral on vertical slices is the original signal. Thus the transform itself is explicitly invertible: we can always get back the original from the transform. Its absolute value may, under some conditions on the windowing function, be inverted as a matter of principle, though an algorithm to do so is rather unwieldy. Similarly, spectral reassignment can be shown to be invertible, as a matter of principle, under certain conditions on the reassignment operator; for instance, instantaneous time-frequency and other complex analytic operators can be inverted, in the sense that all relevant information to permit inversion is preserved; this does not mean, however, that a practical way to do so suggests itself. We shall discuss recent progress on practical algorithms which preserve phase information in reassignment to permit inversion.

3:25

2pSC5. Hidden in plain view: Analysis and synthesis of speech features from derivatives of the short-time phase spectrum. Kelly Fitz (Starkey Hearing Res. Ctr., 2150 Shattuck Ave., Ste. 408, Berkeley, CA 94704, kelly_fitz@starkey.com)

The time-frequency reassigned spectrogram has been shown to yield a clean and readable representation of the distribution of energy in speech signals. In this algorithm, precise, signal-dependent loci for time-frequency analysis data are computed from first-order partial derivatives of short-time spectral phase. In this presentation, evidence for the significance of higher-order derivatives of spectral phase will be presented along with algorithms for computing them. It will be shown that harvesting the information in the short-time phase spectrum allows the construction of an enhanced harmonic model of speech signals, having precise instantaneous frequency data for the harmonic components, and greatly reduced temporal smearing relative to traditional additive models. This enhanced model yields high-quality, artifact-free synthesis of sounds from modified and transformed time-frequency data.

Contributed Paper

3:50

2pSC6. What's wrong with these formants? Sean A. Fulop (Dept. of Linguist., California State Univ. Fresno, PB92 5245 N. Backer Ave., Fresno, CA, 93705, sfulop@csufresno.edu)

Speech is frequently modeled as an all-pole (autoregressive) linear filter applied to a voice source, commonly called Linear Predictive Coding. This model has been used for many years to provide measurements of the resonance frequencies (formants) characterizing speech sounds. That notwithstanding, it has also been well recognized that the use of this method to measure formants is very often fraught with problems, but just how serious are these problems, and can anything be done to improve upon LPC formant measurements? A few examples will demonstrate that the problems plaguing LPC formant measurements are too often intractable, and frequently yield formant values which are nowhere near the truth. For example, the voice bar will be shown to falsely lower the reported first formant value in many simple cases. It might even be suggested that reliance on this technique has impeded our understanding of exotic speech sounds which do not display a typical formant pattern. Precision formant measurements can be provided in most instances by reassigned spectrograms that have been specially processed to highlight signal components. Examples will be shown and used as benchmarks against which the LPC method can at last be properly evaluated.

Session 2pUW

Underwater Acoustics and Acoustical Oceanography: Passive Imaging and Monitoring Using Random Wavefields

Karim G. Sabra, Cochair

Scripps Inst. of Oceanography, Marine Physical Lab., 9500 Gilman Dr., La Jolla, CA 92093-0238

Martin Siderius, Cochair

Heat Light and Sound Research Inc., 12730 High Bluff Dr., Suite 130, San Diego, CA 92130

Chair's Introduction—1:00

Invited Papers

1:05

2pUW1. Travel time estimation from noisy data in random media. George Papanicolaou (Dept. of Mathematics, Stanford Univ., Stanford, CA)

We consider the estimation of Green's functions, or travel times, from noisy data when the medium is not homogeneous. In the homogeneous case travel times can be estimated from cross correlations of noisy traces at the two points between which the travel time is to be estimated, given information about the nature and extent of the noise sources. In a randomly inhomogeneous medium the noisy traces contain information not only about the homogeneous background but also about the inhomogeneities. It is shown how to assess this effect and the limits it puts on travel time estimation.

1:30

2pUW2. Monitoring dynamic matter using the mesoscopic phase statistics of random wave fields. John H. Page, Michael L. Cowan, W. Kurt Hildebrand, Tomohisa Norisuye (Dept. of Phys., Univ. of Manitoba, Winnipeg, MB, Canada R3T 2N2), Domitille Anache-Menier, and Bart A. van Tiggelen (CNRS/Universite Joseph Fourier, Maison des Magisteres, BP 166, 38042 Grenoble, France)

In strongly scattering materials, multiple scattering tends to randomize the phase of transmitted or reflected waves and, as a result, the phase has often been overlooked. In this talk, the use of phase information to monitor the dynamics of multiply scattering media will be described and illustrated through measurements of the temporal fluctuations of ultrasonic waves transmitted through a time-varying mesoscopic sample. The probability distribution of the wrapped phase difference as a function of evolution time, as well as its variance, is measured and compared with theoretical predictions based on circular Gaussian (C1) statistics. Excellent agreement is found. A fundamental relationship between the variance in the phase of the transmitted waves and the fluctuations in the phase of individual scattering paths is predicted theoretically and verified experimentally. This relationship not only gives deeper insight into the physics of the phase of multiply scattered waves, but also provides a new way of probing the motion of the scatterers in the medium. To investigate dynamics on longer time scales, we also investigate the variance and correlations of the cumulative phase. This combination of wrapped and cumulative phase measurements allows both the short and long time dynamics to be probed with excellent sensitivity.

1:55

2pUW3. Energy propagation and localization in disordered media. Kasper van Wijk (Dept. of Geosciences, Boise State Univ., 1910 Univ. Dr., Boise, ID 83712-1536, kasper@cgiss.boisestate.edu) and Matthew Haney (Alaska Volcano Observatory, Anchorage, AK 99508)

While waves in 3D random media still pose quite the challenge to use in model parameter estimation, the complicated paths of multiply scattered waves can be tracked with greater ease in 1D or quasi-1D media. Applying the model of radiative transfer (RT) to the squared waveforms, I will present results of experimental surface-wave multiple scattering, as well as numerical and theoretical aspects of energy propagation in layered media. In the case of surface-wave scattering, it is possible to estimate the average scattering and intrinsic absorption lengths in a laboratory ultrasonic experiment independently. This may be useful in partial saturated materials such as soils and rock, where intrinsic absorption is often related to the movement of fluids in the pore space. Especially in layered media, multiply scattered waves can interfere with one another. An example of strong wave interference is wave localization. Normally, diffusion and RT do not account for wave interference, but new additional terms in RT for layered media explain wave interference and localization: phenomena. Ultimately, the goal is to extend these results for parameter estimation in random media— once understood fully in lower dimensions—to three-dimensional problems.

3100   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3100**

2:20

2pUW4. Extraction of the Green's function from ambient fluctuations for general linear systems. Roel Snieder (Ctr. for Wave Phenomena, Colorado School of Mines, 1500 Illinois Str., Golden, CO 80401-1887, rsnieder@mines.edu) and Kees Wapenaar (Delft Univ. of Technol., 2600 GA Delft, The Netherlands)

The extraction of the Green's function of acoustic and elastic waves from ambient fluctuations is by now a technique that is theoretically well-described and that has successfully been used in different applications. We show theoretically that the principle of the extraction of the Green's function can be generalized to a wide class of linear systems. These new applications include the diffusion equation, Maxwell's equations, a vibrating beam, and the Schröedinger equation. For systems that are invariant for time-reversal it suffices to have sources of ambient fluctuations on a surface that bounds the region of interest. When the invariance for time-reversal is broken; as, for example, in the case of the diffusion equation or for wave propagation in attenuating media, one also needs sources of ambient fluctuations throughout the volume. This work opens up new opportunities to extract the Green's function from ambient fluctuations that include electromagnetic fields in conducting media, flow in porous media, wave propagation in attenuating media, monitoring of mechanical structures, and quantum mechanics.

2:45

2pUW5. Dispersion curves and small-scale geophysics using noise cross-correlation techniques. Philippe Roux, Pierre Gouedard, and Cecile Cornou (LGIT, CNRS 5559, Universite Joseph Fourier, Grenoble, France)

It has been demonstrated, both theoretically and experimentally, that the Green's function between two receivers can be retrieved from cross-correlation of isotropic noise records. Since surface waves dominate noise records in geophysics, tomographic inversion using noise correlation techniques have been performed from Rayleigh waves so far. However, very few numerical studies implying surface waves have been conducted to confirm the extraction of the true dispersion curves from noise correlation in a complicated sedimentary ground model. In this work, synthetic noise has been generated in a small-scale (<1 km) numerical realistic environment and classical processing techniques are applied to retrieve the phase velocity dispersion curves in the medium, first step toward an inversion. We compare results obtained from SPAC (spatial auto-correlation method) and noise correlation techniques on a ten-element array. Two cases are presented in the (1–20 Hz) frequency bandwidth that corresponds to an isotropic or a directional wavefield noise.

3:10

2pUW6. Ambient seismic noise and teleseismic tomography in the western USA: High-resolution 3-D model of the crust and upper mantle from Earthscope/USArray. Yingjie Yang, Michael Ritzwoller, Morgan Moschetti (Ctr. for Imaging the Earths Interior, Dept. of Phys., Univ. of Colorado at Boulder, Boulder, CO 80309), and Donald Forsyth (Brown Univ., Providence, RI 02912)

Short-period surface wave dispersion measurements are extremely hard to obtain from teleseismic events due to scattering and attenuation. Ambient seismic noise is rich in short-period surface waves from which the Rayleigh wave Green function between pairs of stations can be extracted by cross-correlating long noise sequences. Tomography based on surface wave dispersion obtained from the estimated Green functions has been shown to produce high-resolution, short-period (6–30 s) surface wave dispersion maps that principally image crustal geological units (e.g., southern California: Shapiro *et al.*, 2006; Europe: Yang *et al.*, 2007). In this study, we measure phase velocity dispersion curves from the ambient noise cross-correlations to obtain phase velocity maps at periods from 6 to 30 using data from the transportable array component of USArray. A two-plane-wave tomography method including finite-frequency effects was employed to obtain phase velocity maps at complementary periods from 25 to 150 using teleseismic events. The combined phase velocity data set from 6 to 150 is used to invert for high-resolution 3-D Vs structure from the surface to $\sim$ 200 km depth beneath the western USA. The new 3-D Vs model can be used to interpret regional tectonics, model seismic wave propagation, and improve earthquake location.

3:35−3:50   Break

Contributed Papers

3:50

2pUW7. Monitoring a volcano with passive image interferometry. Christoph Sens-Schönfelder[a] and Ulrich Wegler[b] (Universität Leipzig, Talstrasse 35, 04103 Leipzig, Germany)

Ambient seismic noise has been used successfully as a source of information for structural investigations. Ballistic surface as well as body waves were reconstructed by correlation of noise and used in tomographic studies. In these cases it is of course assumed that the medium under study is stationary, i.e., that the reconstructed Green's function does not change with time. In this contribution we show that medium changes can well be monitored by means of subtle changes in the Green's function. Using an interferometric approach applied to the coda part of the Green's function, we detect temporal changes of delay times on Merapi volcano (Indonesia). The changes of delay time depend on lapse time, which indicates that the velocity changes inside the volcano are spatially heterogeneous. We present a hydrological model that can explain the temporal changes of delay time as well as its lapse time dependence changes of the ground water level induced by precipitation. From this analysis we conclude that

(a) seismic coda can be practically retrieved from noise correlations, (b) temporal changes can be monitored with noise correlations, (c) even spatial heterogeneity of the changes can be identified, and (d) the coda retrieved from noise correlations is composed body waves. [a] Currently at LGIT Grenoble, France. [b] Currently at SZGRF/BGR Erlangen, Germany.

4:05

2pUW8. Passive measurements of random wave fields in an instrumented structure. Karl A. Fisher, David H. Chambers, and Sean K. Lehman (Lawrence Livermore Natl. Lab., Livermore, CA 94551)

A passive measurement system using fiber Bragg gratings is presented to interrogate the health of an instrumented part. Estimation of the structures Green's function from diffuse sound fields present during typical operating conditions is the basis for our approach. Experimental studies are conducted using coherent processing techniques of random and generated sound fields to investigate a structure for defects and/or deviations from an initial or pristine state. We are interested in developing a moni-

toring process that is minimally invasive and robust enough to survive the environment that the structure operates in and provide a quantitative assessment of the structure throughout its lifetime.

4:20

2pUW9. Passive *in vivo* elastography from skeletal muscle noise. Karim G. Sabra, Stephane Conti, Philippe Roux, and W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., UC San Diego, San Diego, CA 92093-0238, ksabra@mpl.ucsd.edu)

Measuring the *in vivo* elastic properties of muscles (e.g., stiffness) provides a means for diagnosing and monitoring muscular activity: muscles typically become ''harder'' during contraction occurring through physiological changes. Standard elastography imaging techniques estimate soft tissue (e.g., skeletal muscle, breast) stiffness using propagating shear waves in the human body generated by an external active source (e.g., indentation techniques, ultrasonic radiation force). We demonstrated a passive *in vivo* elastography technique without an active external radiation source. This technique instead uses cross-correlations of contracting skeletal muscle noise recorded with skin-mounted sensors. The coherent arrivals emerge from a correlation process that accumulates contributions over time from noise sources whose propagation paths pass through both sensors successively. Each passive sensor becomes a virtual *in vivo* shear wave source. The results point to a low-cost, noninvasive technique for monitoring biomechanical *in vivo* muscle properties. The efficacy of the passive elastography technique originates from the high density of cross paths between all sensor pairs potentially achieving the same sensitivity obtained from active elastography methods. The application of this passive elastography technique for constructing biomechanical models of *in vivo* muscle properties will be discussed.

4:35

2pUW10. Extracting the local Green's function from ocean noise on a horizontal array. Stephanie E. Fried, Karim G. Sabra, and William A. Kuperman (Scripps Inst. of Oceanogr., 9500 Gilman Dr, La Jolla, CA 92093-0238, sefried@ucsd.edu)

Time delays, associated with different ray paths between the elements of a bottom hydrophone array, can be extracted using ambient noise cross-correlations [Sabra *et al.*, J. Acoust. Soc. Am. **114**, 2462 (2003)]. This is confirmed using long time noise recordings that were collected in May 1995 near the S. California coast at an average depth of 21 m. The noise is mainly biological in the frequency range of 250–700 Hz [D. Spain *et al.*, J. Acoust. Soc. Am. **99**, 2453 (1996)]. The cross-correlations of noise from a horizontal array can be used to extract the local Green's function from the ocean ambient noise. [Research supported by ONR.]

4:50

2pUW11. Large-area acoustic field characterization in shallow water by using broadband ship noise measured on the New Jersey Shelf. Altan Turgut (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

Acoustic signals emitted by ships of opportunity (merchant ships) are simultaneously recorded on three vertical line arrays (VLAs) during the New Jersey Shelf RAGS03 experiment. Although the single-receiver correlation between the VLAs (separated by ∼ 10 km) was very low, a Bartlett correlator between each VLA produced well-defined striation patterns. Waveguide invariant theory has been applied to Bartlett correlator output to obtain range ratios of the noise source to the receiver arrays [A. Turgut, J. Acoust. Soc. Am. **118**, 1857 (2005)]. Trajectories of striations are identified by the Hough transform that converts a difficult global detection problem in the image domain into a simpler local peak detection problem. The striation patterns are also observed when the Bartlett correlator is applied to different time segments of the noise data. This provided an opportunity of obtaining reference acoustic field data for winter conditions if the location of the noise source is known *a priori*. Analysis of the RAGS03 data indicates the feasibility of source localization and acoustic field characterization by using broadband noise signals emitted by distant surface ships. [Work supported by ONR.]

5:05

2pUW12. Green's function retrieval through ocean acoustic interferometry. Laura A. Brooks[a] and Peter Gerstoft (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA)

A stationary phase argument is used to describe the relationship between the stacked cross-correlations from a line of vertical sources, located in the same vertical plane as two receivers, and the Green's function between said receivers. Results and simulations which demonstrate the approach and are in agreement with those of a modal based approach used by others are presented. Results indicate that the stacked cross-correlations can be directly related to the shaded Green's function, so long as the modal continuum of any sediment layers is negligible. Preliminary results from the SWO6 experiments are discussed. [Work supported by ONR, Fulbright (sponsored by Clough Engineering), and DSTO Australia.] [a] Also at the School of Mechanical Engineering, University of Adelaide, Australia.

5:20

2pUW13. Environmental effects on passive fathometry and bottom characterization. Steven L. Means (Naval Res. Lab., Code 7120, 4555 Overlook Ave. SW, Washington, DC 20375, means@wave.nrl.navy.mil) and Martin Siderius (HLS Res. Inc., San Diego, CA 92130)

Recently, Siderius, Harrison, and Porter developed a method, based on the Fourier synthesis of the cross-spectral density from nearby positions, to exploit the oceans' coherent ambient noise field due to breaking waves to make measurements of the bottom and subbottom properties. During 2006 breaking wave noise measurements were made in the shallow waters (25 m) approximately 75 km off the coast of Savannah, GA on a 32-phone, three nested-aperture, vertical hydrophone array that was deployed 100 m from a Navy tower that stands 50 m above the water surface. The Skidaway Institute of Oceanography operates a suite of instruments for measurements of both atmospheric and oceanic conditions at the tower. Data were collected in a variety of environmental conditions, with wind speeds ranging from 5–21 m/s and wave heights of 1–3.4 m. The data is analyzed to quantify the performance of the passive fathometer methods as a function of the wind speeds, wave conditions, and averaging times. The results will be compared with ground-truth measurements made at the tower site prior to its construction. [Work supported by ONR base funding at NRL.]

5:35

2pUW14. Array limitations of coherent noise processing for geoacoustic inversion. Peter Gerstoft, Chen-Fen Huang, William S. Hodgkiss (Marine Physical Lab., Univ. of California San Diego, CA 92093-0238), and Martin Siderius (Heat Light and Sound Res. Inc, CA 92130)

Ocean acoustic noise can efficiently be processed to extract Greens function information from noise [Roux *et al.*, JASA 2004, Siderius *et al.*, JASA 2006]. By crosscorrelating the ambient noise field from two sensors, it is possible to extract the impulse response between the two sensors including bottom and sub-bottom bounces if the noise field is homogenous. Since the major part of the ocean-noise is horizontal and thus strongly inhomogeneous, this processing does not extract the vertical bouncing energy well. Using array processing it is possible extracting the vertical bouncing energy. However, this beamforming-like approach can just be done for frequencies less than the array design frequency. For higher frequencies the side lobes in the beam pattern will make the horizontal energy appear as vertical and can thus not be used. Vertical velocity sensors can filter out some of this horizontal energy and a vertical velocity sensor array can be used at higher frequencies than a pressure sensor array. When this noise processing is used on a vertical array, it gives valuable information about the subbottom near the array.

Meeting of Accredited Standards Committee (ASC) S12 Noise

R. D. Hellweg, Chair, S12

Hewlett Packard Co., Acoustics Lab, MR01-2/K15, 200 Forest Street, Marlborough, MA 01752

W. J. Murphy, Vice Chair, S12

NIOSH, 4676 Columbia Parkway, Mail Stop C27, Cincinnati, OH 45226

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 10:30 a.m. on Tuesday, 5 June 2007.

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.

2p TUE. PM

NOTE: Session attendees should enter the LDS Conference Center through door 13 on the southwest corner of the building. You will be directed to the Media Room by the Center's hosting personnel

Session 3aAA

Architectural Acoustics and Musical Acoustics: Effects of Rooms on the Sound of Organs

Bruce C. Olson, Chair

Olson Sound Design, 8717 Humboldt Ave., North, Brooklyn Park, MN 55444-1320

Chair's Introduction—8:15

Contributed Paper

8:20

3aAA1. A qualitative criterion for the reverberant fields for organ music and a means of obtaining it. Bertram Y. Kinzey, Jr (Dept. of Architecture, Univ. of Florida) and Michael Ermann (Virginia Tech School of Architecture, Blacksburg, VA 24060)

Organists and organ builders desire long reverberation times for their instruments. A simple statement of a time is an inadequate specification for the acoustic response of a venue intended for medium to larger organs. An important added specification requirement is that the reverberation decay curve be steep at first and be much less so later if satisfactory clarity is to be achieved. Data is presented to illustrate and establish this criterion. Research is cited to demonstrate how the double slope reverberation decay can be achieved by the use of coupled spaces adjacent to a main performing room. Data is given to demonstrate how some symphony halls have realized the use of this technique.

Invited Papers

8:35

3aAA2. Church and concert hall acoustics. Charles Hendrickson (1403 N. 5th St., St. Peter, MN 56082)

The separate acoustical criteria for churches and concert halls are presented from the viewpoint of a pipe organ builder—a former college physics teacher and president of the Associated Pipe Organ Builders of America. Developed from considerations of the different functions and purposes of churches and concert halls, the criteria are markedly different and even contradictory for these situations. Beranek's ''Music, Acoustics & Architecture'' is referenced.

9:05

3aAA3. Cathedral acoustics and organ sound reinforcement. Peter Mapp (Peter Mapp + Assoc., 5 Worthington Way, Colchester, CO3 4JZ, UK)

The acoustic characteristics of 4 UK cathedrals have been investigated. In three of the buildings it is reported that the organ requires additional support or reinforcement to cope with the large volume of the space involved. All 4 buildings were found to exhibit multiple coupled spaces making it difficult to accurately assess the active acoustic volume. In order help to overcome the perceived lack of 'volume' from the organ, in two of the cathedrals, it is proposed to install a second rank of pipes some distance down the nave from the main organ. The acoustic and psychoacoustic implications if this approach are discussed. An alternative means of providing support for the organ via the installed sound reinforcement system has also been investigated and the electroacoustic and subjective requirements for this are discussed and reviewed.

9:35

3aAA4. Clarifying reverberation for organ performance. Jack M. Bethards (Schoenstein & Co. Organ Builders of San Francisco, 4001 Industrial Way, Benicia, CA 94510, jack@schoenstein.com)

In the conceptual phase of design or remodeling of a space for organ performance, the acoustical consultant will have a limited time in which to assess the potential for reverberation and to guide the designer in maximizing this potential. In a place of worship, the organ is likely to serve as an accompaniment to the choir and also as a solo performance instrument, both of which have their distinct acoustical requirements. It is important for the designer to fully understand the intended uses of the organ and the qualities that the organ builder wishes to emphasize. While attaining sufficient reverberation for organ performance is often a challenging task, too much reverberation, which can blur form, harmonic structure, rhythm articulation, and dynamic contrasts, must also be avoided. The architectural factors to be considered in this careful balance are reviewed together with examples of successful organ installations.

10:20

3aAA5. The behavior of a pipe organ in a very large room: An organist's perspective of the Latter Day Saints (LDS) Conference Center organ. John Longhurst (Mormon Tabernacle Choir and Orchestra at Temple Square, Salt Lake City, UT 84150)

The LDS Conference Center in Salt Lake City houses one of the largest theater-style auditoriums in the world. The hall seats 21,000 and incorporates a prominent pipe organ built by Schoenstein & Co. of San Francisco. The organ is characteristic of the American Romantic style, while employing a symphonic tonal approach with the richness and warmth of English instruments. A five-manual console controls the 7,708 pipes of its 130 ranks, which are spread across seven divisions. This presentation will focus on the rationale for its tonal design, dictated in part by the acoustics of the hall, and provide live demonstrations of its capabilities.

WEDNESDAY MORNING, 6 JUNE 2007

SALON B, 8:30 TO 10:30 A.M.

Session 3aAB

Animal Bioacoustics: Transmission, Detection and Analysis of Bioacoustic Signals

Whitlow W. L. Au, Chair

Hawaii Inst. of Marine Biology, P. O. Box 1106, Kailua, HI 96734

Contributed Papers

8:30

3aAB1. Localization of killer whale (*Orcinus orca*) vocalizations using a triangular hydrophone array in Johnstone Strait, British Columbia, Canada. Dawn M. Grebner, David L. Bradley, and Dean E. Capone (Penn State Univ., State College, PA 16804)

A triangular hydrophone array was bottom-mounted in Johnstone Strait, in summer 2006, to localize sounds emitted by individual or small groups of orcas. Spatial and temporal positioning of individual orcas was obtained from a cliff-site observation platform; a video camera recorded surface behaviors and associations while a theodolite was used to obtain vertical and azimuthal angles needed to calculate distances to the animals. The time of arrival differences of the orca sounds at each hydrophone was used to determine the angle of arrival of the sound. Sounds were then aligned in time and space with the orcas' corresponding movement tracks. The ability to spatially and temporally localize the vocalizations and behaviors of an isolated orca provides insight into how an individual or group of individuals may function within a matrilineal unit or pod. The objective of this study is to determine if there are defined "roles" in the vocal behaviors of individuals (adult males, adult females, juveniles, calves) within small groups of orcas; specifically, individual differences in orca calling behavior that regulates group movement and changes in group behavior. Results from analysis of pod encounters will be discussed and correlation of vocalizations and specific individual behavior addressed.

8:45

3aAB2. Evaluation of an acoustical-optical platform for measuring species specific abundance in a weir trap. Jennifer L. Miksis-Olds (Univ. of Massachusetts, School for Marine Sci. and Technol., 838 S. Rodney French Blvd., New Bedford, MA 02744, jmiksis@umassd.edu)

Data recorded by a bottom mounted survey platform deployed within the opening of a fishing weir were used to calculate absolute abundance estimates for comparison to the weir catch. Species specific abundance estimates were calculated from a combination of sonar and video information recorded by the acoustical-optical platform (AOP). Individual targets were detected in the sonar images using an echo counting technique. Video images were used to provide species identification of acoustic targets within an approximate 10 m range. The number and identification of acoustic targets detected during each deployment was positively correlated with the total catch weights and dominant species in each deployment. The algorithm used to calculate AOP abundance estimates tended to underes-

timate abundance when more than 50 individuals of the same species were caught and overestimate abundance when less than 50 individuals were caught. Target strength, bearing, and range distributions for individual targets are presented in addition to the species specific percent difference between the AOP estimates and absolute abundances of the weir catch. The agreement between the species specific AOP estimates and absolute abundances of each species indicate that the AOP has the potential to provide data for accurate estimates of multispecies abundance. [Work supported by NOAA.]

9:00

3aAB3. Underwater vocalizations of male harbor seals (*Phoca vitulina*) during the mating season at Brookfield Zoo. Mehgan Clark and Jeanette Thomas (Dept. of Biol. Sci. WIU, 1 Univ. Circle, Macomb, IL 61455)

Underwater vocalizations of harbor seals were recorded between April and October of 2006 in many settings, including male only, two females prior to the breeding season, and all seals as a group before, during, and after the breeding season. Recordings were made using a Dell XPS M140 laptop, with a linear frequency response up to 96 kHz, and Goldwave sound recording software. All recordings were analyzed using Spectro-gram 14.0 software. Three call types were identified: short roars, long roars, and growls. A total of 736 roars were identified and 256 randomly sampled roars were analyzed. A total of 109 growls were identified and analyzed. The frequency range of the call types was 209 to 3406 Hz, based on dominant minimum and dominant maximum frequency measurements, and duration ranged from 0.8 to 10.5 s. The male produced stereotypical roars in late June. As there was only one male in the group, it is suggested that the function of this roar is to announce breeding condition or to attract a female for mating, not for competition among males.

9:15

3aAB4. Nocturnal activity by transient killer whales at St. Paul Island, AK. Kelly Newman and Alan Springer (School of Fisheries and Ocean Sci., Univ. of Alaska Fairbanks, 245 O'Neill Blvd., Fairbanks, AK 99775, k.newman@sfos.uaf.edu)

Observations of killer whales in the act of predation are rare because feeding bouts are infrequent in relation to sighting opportunities, and they apparently feed at night as well as during the day. A pop-up autonomous recording unit (ARU) was used to expand the observation window of

transient (mammal-eating) killer whales by monitoring them continuously from 22 June to 12 July 2006 near a northern fur seal rookery at St. Paul Island, AK, a predation hot-spot in the Bering Sea. The ARU recorded at a sampling rate of 0.5–16 000 Hz. Killer whale vocalizations were detected on 19 of 20 days. Most vocalizations occurred from the end of civil twilight (00:26 on 1 July) through mid-morning. Fewest vocalizations occurred between 1800 and 2400. Vocal whales were most likely feeding, based on the paradigm that transients are mostly silent except when attacking prey. Call types are being identified and vocalizations are being correlated with visual observations and fur seal activity at the rookery. Nocturnal hunting is likely one important reason why it has been difficult to understand the foraging behavior of these elusive predators and must be considered when drawing conclusions about the roles transient killer whales play in marine ecosystems.

9:30–9:45   Break

9:45

3aAB5. Implications of shallow water propagation for ranging by marine mammals. Eduardo Mercado III, Jennifer N. Schneider, Sean R. Green, and Rachael D. Rubin   (Dept. of Psych., Univ. at Buffalo, the State Univ. of New York, Buffalo, NY 14260)

Marine mammals rely heavily on acoustic communication to locate one another in low-visibility underwater environments. How these animals localize sounds sources is poorly understood. The speed of sound, attenuation, and reverberation in shallow water environments differ from those in terrestrial environments such that intensity cues alone may not be sufficient for ranging. The purpose of this study was to examine sound propagation in shallow, coastal waters near Rincon, Puerto Rico, and to determine what acoustic cues provide consistent information about source distance. Narrow-band tones and broadband sounds of similar frequency content (100–22 000 Hz) and duration (2.0 s) to sound units of humpback whale song were broadcast from an underwater speaker. Sounds were recorded at known distances ranging from 70–850 m from the sound source. The received intensity was negatively correlated with distance propagated; broadband sounds showed a stronger correlation than narrow-band sounds. Analysis of the intensity-normalized spectral content of received sounds revealed a complex pattern of sound degradation: the distribution of peaks and notches within the frequency band alternating systematically as a function of distance. These results indicate that intensity cues to source distance are frequency dependent and that marine mammals may use cues other than intensity when ranging. [This work was supported in part by the Earthwatch Foundation and the UB Honors Program.]

10:00

3aAB6. Acoustic backscatter from a diving dolphin. Whitlow W. L. Au   (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Dorian S. Houser   (Biomimetica, Santee, CA 92071), and Lois A. Dankiewicz   (Science Applications Intl. Corp., San Diego, CA 92110)

The acoustic backscatter of a diving Atlantic bottlenose dolphin (*Tursiops truncatus*) was measured using Simrad EK-60 echosounders operating at 70, 120, and 200 kHz. The animal was trained to dive to a small platform suspended with a nylon rope from a small boat to various depths, with the maximum depth being 100 m. When the dolphin reached the platform, it was required to station on a bite plate attached to the platform until an acoustic signal was sent to release the animal. A video camera on the platform with the signal sent up on a separate line was used to monitor the animal on the bite plate. A time-depth recorder was also attached to the dolphin to determine the average rate of decent and ascent. Echosounding records were collected while the dolphin descended to and ascended from the platform. A total of 15 echosounding records with the dolphin swimming within the beam of the transducers were collected. Preliminary analysis of the data indicated that target strength varied from about -26 dB to -40 dB with a slight frequency dependency. The echoes from the dolphin were structured with the maximum echoes originating from the vicinity of the lungs.

10:15

3aAB7. Classification of wolf call types using remote sensor technology. Deborah Curless, Marie A. Roch, Shyam Madhusudhana (Dept. of Computer Sci., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182-7720), Melissa S. Soldevilla, John A. Hildebrand (Univ. of California at San Diego, La Jolla, CA 92093-0205), Elizabeth Baker, Dan Moriarty   (Univ. of San Diego, San Diego, CA, 92110-2492), and Kim Miller   (California Wolf Ctr., Julian, CA 92036)

A system is presented for the near-real-time classification of calls from captive Alaskan (*Canis lupus occidentalis*) and Mexican (*Canis lupus baileyi*) wolves as well as other ambient sounds such as bird calls and anthropomorphic noise. Signals are detected on a Viper network node at The California Wolf Center using a signal-to-noise ratio based call activity detector. These calls are transmitted over a regional high-speed wireless network (HPWREN) to a remote processing facility located 50 miles from the collection site. Cepstral feature vectors are extracted from candidate calls at the remote facility and are classified using a hidden Markov model pattern recognition system trained with manually labeled data representing the wolf ethogram and other common noises in the soundscape. In an offline evaluation with known segmented data, the system achieves an 84% accuracy. [This work is supported by HPWREN, NSF Grant No. 0426879.]

Session 3aAO

Acoustical Oceanography, Underwater Acoustics, Engineering Acoustics, and Signal Processing in Acoustics: Acoustic Sensing of the Ocean and Seabed Using Gliders and Autonomous Underwater Vehicles

James F. Lynch, Chair

Woods Hole Oceanographic Inst., MS 11, Woods Hole, MA 02543-1541

Chair's Introduction—8:00

Invited Paper

8:05

3aAO1. Autonomous undersea vehicles and acoustical oceanography. Thomas Austin and Ben Allen (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Autonomous undersea vehicles (AUVs) have been actively increasing in their capabilities over the past 10 years. Recent developments in the academic sector have transitioned to become commercially available technologies which are now available for general use in academic, commercial and military applications. The Oceanographic Systems Laboratory at Woods Hole Oceanographic Institution has been a leader in the development of robotic vehicles, with most development funding coming from the Office of Naval Research. The result of this development effort has been the REMUS family of AUV systems. Originally designed primarily for shallow water hydrographic reconnaissance and mine-hunting, these systems have since been adapted and utilized for many other applications including deep ocean search and survey, physical oceanography, and more recently acoustical oceanography. New, larger systems contain sufficient energy and payload capabilities to tow long hydrophone arrays, and can carry large projectors allowing active and passive techniques to be used in unmanned, un-tethered systems at any depth in the oceans. Recent developments in these unmanned vehicle systems will be discussed along with new payload capabilities now under development for acoustical oceanography.

9:05−9:20   Break

Contributed Papers

9:20

3aAO2. Underwater acoustic measurements with a flying wing glider. Gerald L. D'Spain, Richard Zimmerman, Scott A. Jenkins (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 93940-0701), James C. Luby, and Peter Brodsky (Univ. of Washington, Seattle, WA 98105)

Liberdade, a new class of underwater glider based on a flying wing design, has been under development for the past 3 years in a joint project between the Marine Physical Laboratory, Scripps Institution of Oceanography and the Applied Physics Laboratory, University of Washington. This hydrodynamically efficient design maximizes the horizontal distance traveled between changes in buoyancy, thereby minimizing average power consumed in horizontal transport to achieve ''persistence.'' The first fully autonomous glider of this class, ''XRay,'' was deployed and operated successfully in the Monterey Bay 2006 experiment. Communications, including real-time glider status reports, were accomplished using an underwater acoustic modem as well as with an Iridium satellite system while on the surface. The payload included hydrophone array, with 10 kHz per channel bandwidth, located in a sonar dome along the leading edge of the 6.1-m-span wing. Narrowband tones from 3.0 to 8.5 kHz were transmitted from a ship-deployed controlled underwater source. During the glider's flight, lift-to-drag ratios (equal to the inverse of the glide slope) exceeded 10/1. However, specific flight behaviors that deviated from this efficient horizontal transport mode allowed for improved detection and localization by the hydrophone array. [Work sponsored by the Office of Naval Research.]

9:35

3aAO3. High-resolution repeat bathymetric mapping using a mid-size autonomous underwater vehicle. Gerald L. D'Spain, Richard Zimmerman, C. David Chadwell, Mark Zumberge, John Blum, John Orcutt, Neal Driscoll, Jeff Dingler, Graham Kent (Scripps Inst. of Oceanogr., La Jolla, CA 93940), Hugh H. Banon, Graham Openshaw, and Philippe Jeanjean (BP America, Houston, TX)

Shipborne bathymetric surveys south of Santa Barbara, CA indicate that massive slope failures have occurred along the northern flank of the Santa Barbara Basin. This region is seismically active and has the highest sedimentation rates along the California coast due to rapid erosion of the Transverse Ranges. A set of geodetic monitoring systems is being developed to study the character and deformation of the ocean bottom in this region, particularly that of a prominent east-west trending crack aligned with the head scarp of one of the underwater landslides. One component is a high-resolution repeat bathymetric mapping system composed of a multibeam sonar mounted inside a Bluefin 21 AUV. Navigation to the decimeter level is provided by an inertial navigation system and accurate depth sensor, and by a long baseline (LBL) acoustic transponder system previously developed to monitor tectonic plate motions. The quality of the data, particularly that from the high-precision LBL system, is enhanced significantly by the decrease in the radiated acoustic and vibration noise of the AUV to levels at, or below, ocean background noise levels through modifications to the AUV propulsion system. [Work sponsored by British Petroleum and the Office of Naval Research.]

3107   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   3107

3a WED. AM

3aAO4. Applications of an ultra low-drag towed array deployed from a glider. Paul Hursky, Martin Siderius, Michael B. Porter (Heat, Light, and Sound Res. Inc, 12730 High Bluff Dr. Ste. 130, San Diego, CA 92130), Vincent K. McDonald, Mark Gillcrist, Brian Granger, Ryan Jones (Space and Naval Warfare Systems Ctr., San Diego, CA 92152), Aaron Bratten, Andy Huizinga, Peter T. Sullivan, and Susan G. Briest (Space and Naval Warfare Systems Ctr., San Diego, CA 92152)

We have deployed a Webb Research Slocum glider towing a 21-meter 16-element line array in two experiments, the Makai experiment off of Kauai, and the PLUSNet experiment in Monterey Bay. A glider platform provides some unique capabilities. It is very quiet, by virtue of its buoyancy-driven propulsion, and it visits or samples most of the water column on a regular basis, following its sawtooth dive profile. We have collected data in several configurations, including a stop-and-drift mode, in which the array assumes a vertical posture, recording narrowband and broadband controlled sources and ambient noise. We will present the results of processing these datasets, and discuss their implications for surveillance, seabed mapping, and ambient noise measurements.

3aAO5. Acoustic measurements in shallow water using an ocean glider. Georges A. Dossot, James H. Miller, Gopu R. Potty, Kristy A. Moore (Univ. of Rhode Island, Dept. of Ocean Eng., Narragansett Bay Campus, Narragansett, RI 02882), Jason D. Holmes (BBN Technologies, Cambridge, MA 02138), and James F. Lynch (Woods Hole Oceanograph. Inst., Falmouth, MA 02543)

Acoustic data was collected on a single hydrophone attached to a Webb Slocum glider deployed by Rutgers University during the Shallow Water Experiment (SW06) on the continental shelf off New Jersey. The geometry of the experiment provided for adequate recording of the 224 and 400 Hz WHOI tomography sources. The glider periodically surfaced for GPS fixes and data transfer via satellite phone. A synthetic aperture is created through coherent processing of the acoustic data as the glider traveled through the water. A number of issues including varying depth and nonsteady motion must be accounted for due to the trajectory of the glider. However, the glider provides a low-noise and low-speed platform, potentially enabling detection of low level signals. Initial results show that the phase of the tomography signals is linear as the glider traverses the sound field. An acoustic normal mode representation of the field provides the basis for processing of the data similar to Hankel transform approach of Frisk and Lynch [J. Acoust. Soc Am. **76**, 205–216 (1984)] and Holmes, Carey, and Lynch [J. Acoust. Soc. Am. **119**, 3346 (2006)]. Effects of spatial variations in sound speed are examined. [Work sponsored by the Office of Naval Research.]

3aAO6. Graph-theoretic sonar detection performance of maneuverable vehicles in variable propagation environments. Igal Bilik and Jeffrey Krolik (Duke Univ., Durham, NC 27708-0291)

This work addresses the problem of passive acoustic target detection using a maneuverable unmanned underwater vehicle (UUV) equipped with a towed line array. The objective is to assess target detection performance as a function of vehicle maneuverability and environmental variability. Detection performance is evaluated by mapping the search region onto a graph with nodes that constrain the vehicle trajectory and edge weights determined by the local detection probability. Vehicle maneuverability is constrained by limiting the maximum turn angle at each node. The UUV trajectory that maximizes the probability of at least one target detection can then be determined using the computationally efficient Dijkstra algorithm. Increased maneuverability improves detection performance given known environmental variability since the optimal trajectory is able to visit more locations with favorable environmental conditions. The optimal detection performance of a maneuverable array with relatively short array lengths is compared to that of a nonmaneuverable vehicle with a longer array. It is shown that maneuverability can be even more advantageous than array length given sufficiently variable environmental conditions. The impact of uncertainty and/or mismatch of the assumed environmental variability is also studied as a function of vehicle maneuverability and towed-array length. [This work support by ONR 321SP.]

3aAO7. A proposed technique for source localization using an autonomous underwater vehicle. Kevin Cockrell and Henrik Schmidt (Dept. of Mech. Eng, MIT, COE, Cambridge, MA 02139)

It has been of recent interest to use autonomous underwater vehicles (AUVs) with towed arrays for source localization in ocean acoustic waveguides with uncertain environmental parameters. However, AUVs cannot implement source localization techniques designed to be robust against environmental mismatch because such techniques require more computational power than is typically available on an AUV. To address this issue, a technique is proposed that increases robustness to environmental mismatch by utilizing computational resources available before the measurement of the acoustic field is made (i.e., computational resources on the ship from which the AUV is launched). This is done by using a numerical ocean acoustic model to analyze how uncertainty in environmental parameters affects the acoustic pressure levels at the hydrophones of the array. The results of this analysis can then be stored on the AUV's computer and used by the source localization algorithm. The relationship between the proposed technique and matched mode processing is discussed.

Session 3aBBa

Biomedical Ultrasound/Bioresponse to Vibration: Biomedical Applications of Acoustic
Radiation Force Imaging

Mostafa Fatemi, Chair

Mayo Clinic, Dept. of Physiology and Biophysics, 200 First St., SW, Rochester, MN 55905

Chair's Introduction—8:00

Invited Paper

8:05

3aBBa1. Physical and historical bases of biomedical applications of radiation force. Armen Sarvazyan (Artann Labs., 1459 Lower Ferry Rd., West Trenton, NJ 08618)

Radiation force is a universal phenomenon in any wave motion, electromagnetic or acoustic. Radiation force is produced by a change in the density of energy of the propagating wave. The concept of radiation pressure follows from Maxwell's equations (1871), while the idea of electromagnetic wave (light) pressure can be traced back to Johannes Kepler (1619). Acoustic radiation force was first experimentally demonstrated by August Kundt (1874) and its physical bases were analyzed in classical works of Lord Rayleigh (1902), L. Brillouin, and P. Langevin. Until recently, main biomedical application of acoustic radiation force was in measuring acoustic power of therapeutic devices, but during the last 15 years it became one of the hottest areas of biomedical ultrasonics with numerous new applications ranging from acoustical tweezers, targeted drug and gene delivery, manipulating of cells in suspension, increasing the sensitivity of biosensors and immunochemical tests, assessing viscoelastic properties of fluids and biological tissue, and monitoring lesions during therapy. The widest new area of applications of radiation force is related to medical imaging in general and elasticity imaging specifically. Remote excitation of shear stress in a given site in the body using radiation force of focused ultrasound offers numerous new possibilities for medical diagnostics.

Contributed Papers

8:30

3aBBa2. Axial radiation force of a Bessel beam on a sphere, direction reversal of the force, and solid sphere examples. Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

The exact solution for the axial radiation force on an isotropic sphere centered on an acoustic Bessel beam in an inviscid fluid was recently given [P. L. Marston, J. Acoust. Soc. Am. **120**, 3518–3524 (2006)]. Following a review of the physical relevance of such radiation force calculations for acoustic plane waves, the Bessel beam radiation force derivation and results will be summarized. In the ordinary case, the direction of the axial force is parallel to the propagation direction of the beam. Among the predictions given in the publication is the possibility of an axial radiation force on certain fluid spheres for which the force is directed opposite to the propagation direction. In a recent extension of the published work, conditions have been identified for such a force reversal on solid spheres and elastic shells. These calculations give insight into situations where the radiation force of acoustic beams may be used to pull (instead of push) spherical objects. [Research supported in part by NASA.]

8:45

3aBBa3. Harmonic motion detection in a vibrating scattering medium. Matthew W. Urban and James F. Greenleaf (Dept. of Physiol. and Biomed. Eng., Mayo Clinic College of Medicine, 200 First St., SW, Rochester, MN 55905)

Dynamic ultrasound radiation force can induce local tissue vibration. To understand the tissue response, we developed a method to ultrasonically measure the induced motion. We present a model for motion detection in a vibrating scattering medium. We developed a harmonic pulsed excitation method using tonebursts repeated at frequency f_r. This excitation method creates a multifrequency force with components at harmonics of f_r. We use pulse-echo ultrasound and narrowband Kalman filtering to

perform multifrequency motion detection. In an experiment, a 3.0 MHz air-backed transducer was used to create the radiation force, and with the same transducer performed pulse-echo motion detection at 9.0 MHz. Bovine muscle was used for radiation force excitation and motion detection. We present simulation results for a parameterized model of motion detection in a scattering medium. We show experimental measurements of vibration displacement and phase for motion with amplitude of $0.1–10 \ \mu$m. Our simulation model provides a platform to optimize detection of small harmonic motion in a scattering medium. Multifrequency harmonic pulsed excitation and detection methods are presented that induce and measure very small amplitude harmonic motion.

9:00

3aBBa4. Focused ultrasound surgery monitoring using local harmonic motion imaging. Laura Curiel, Rajiv Chopra, and Kullervo Hynynen (Imaging Res., Sunnybrook Health Sci. Ctr., 2075 Bayview Ave., Toronto, ON, M4N 3M5, Canada)

The present study established the feasibility of a real-time monitoring technique for focused ultrasound (FUS) lesion formation using localized harmonic motion imaging. This oscillatory motion was generated within tissues by periodically induced radiation force using a FUS transducer (oscillatory motion frequencies between 50 and 300 Hz). The harmonic motion was estimated using cross correlation of rf ultrasonic signals acquired at different instances during the motion by using a separate US diagnostic transducer excited by a pulser/receiver. The technique was evaluated on rabbit muscle under *in vitro* and *in vivo* conditions. Fourteen FUS lesions were induced *in vivo* inside an MR scanner to obtain simultaneous ultrasound harmonic motion tracking and MR thermometry. The calculated maximum amplitude of the induced harmonic motion before and after the lesion formation was significantly different for all the tested motion frequencies. During the FUS exposure a drop in the maximum amplitude value was observed and could be associated to a lesion forma-

3109   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3109**

3a WED. AM

tion. After the lesion formation, the harmonic motion was measured across the exposure region, and the lesion was detected as a reduction in the maximum motion amplitude value at the coagulated region. [Work supported by NIH Grant R21/R33 CA102884-01.]

9:15–9:30   Break

9:30

3aBBa5. Generation of shear motions by radiation force in a nonlinearly distorted ultrasonic beam. Lev Ostrovsky (Zel Technologies/Univ. of Colorado, 325 Broadway, Boulder, CO 80305, Lev.A.Ostrovsky@noaa.gov)

A new possible mechanism for shear wave generation in biological tissues by focused ultrasound is described. In our previous work it was shown that, besides the classical, dissipative mechanism for this effect, medium inhomogeneity can result in a non-dissipative generation of shear displacements. In this presentation, it is shown that even in a homogeneous water-like medium such as tissue, nonlinear distortions (i.e., harmonic generation) of a focused ultrasonic wave can generate shear motions even without dissipation (i.e., prior to the shock wave formation). Estimates made for biological tissues show that the effect can be significant.

9:45

3aBBa6. Exact solutions of the compressible inverse elasticity problem. Paul E. Barbone (College of Eng., Boston Univ., 110 Cummington St., Boston, MA 02215) and Assad A. Oberai (Rensselaer Polytechnic Inst., Troy, NY 12180)

In elastic modulus imaging or quantitative elastography, tissue stiffness is inferred from an (ultrasonically) measured displacement field. Doing so requires the solution of an elastic inverse problem. We present several exact solutions of the compressible inverse elasticity problem: Given the (possibly transient) dispalcement field measured everywhere in an isotropic, compressible, linear elastic solid, and given density ρ, determine the Lamé parameters $\lambda(x)$ and $\mu(x)$. The cases we treat are: (a) for $\mu(x)$ known *a priori*, find λ; (b) for $\lambda(x)$ known *a priori*, find μ; (c) neither $\lambda(x)$ nor μ is known *a priori*, but Poisson's ratio ν is known; (d) neither $\lambda(x)$ nor μ nor ν is known *a priori*. We present several example applications including 2D and 3D problems, quasistatic and time dependent displacement fields. Many of the results here are drawn from P.E. Barbone and A.A. Oberai ["Elastic Modulus Imaging: Some exact solutions of the compressible elastography inverse problem," Phys. Med. Biol. (in press)].

10:00

3aBBa7. Progress in breast vibro-acoustography: Imaging breast lesions. Azra Alizad, Dana H. Whaley, James F. Greenleaf, and Mostafa Fatemi (Mayo Clinic College of Medicine, Rochester, MN 55905, aza@mayo.edu)

Vibro-acoustography is an imaging method based on the sound emitted due to the harmonic motion of tissue induced by the radiation force of focused ultrasound. This paper presents recent progress in vibro-acoustography of human breast. Matching vibro-acoustography and mammography images are acquired using a combined mammography/vibro-acoustography machine. Images of various types of breast masses are investigated and their characteristics features are compared. The study includes malignant masses, benign masses such as fibroadenomas, cysts, calcified lesions, and surgical scars. Image features, such as lesion borders, interior texture, and speculation pattern, are studied. Vibro-acoustography images demonstrate that calcified masses normally show sharp borders, benign noncalcified masses have smooth texture, and, in some cases, malignant masses are characterized with spiculation. These results, which are based on the data collected from an ongoing study, indicate that the above-mentioned image features may be used for differentiation of breast masses. It is noted that more data from a larger population are needed to establish the results in a statistical sense. [Work supported by NIH Grant EB-00535 and Grant BCTR0504550 from the Susan G. Komen Breast Cancer Foundation. Disclosure: Parts of the techniques used here are patented by MF and JFG.]

10:15

3aBBa8. A noninvasive resonant mode imaging. Xiaoming Zhang, Mostafa Fatemi, Randall R. Kinnick, and James F. Greenleaf (Mayo Clinic, 200 1st St. SW, Rochester, MN 55905)

The mechanical response of tissues to external forces has gained considerable interest in medical diagnosis. One approach of imaging the tissue elastic properties is to apply an external force on the surface of the body. Another approach is to generate a localized force inside the tissue with the radiation force of ultrasound. In this paper, a noninvasive method for imaging the vibration mode of an object is investigated. The radiation force of ultrasound is used to scan the object at a resonant frequency of the object. The vibration of the object is measured by laser and the resulting acoustic emission from the object is measured by a hydrophone. Experimental studies are carried out on a mechanical heart valve and arterial phantoms. The mode images on the valve are made by the hydrophone measurement and confirmed by FEM simulations. Compared with conventional B-scan imaging on arterial phantoms, the mode imaging can show not only the interface of the artery and the gelatin, but also the vibration modes of the artery. The images taken on the phantom surface suggest that an image of an interior artery can be made by vibration measurements on the surface of the body.

Session 3aBBb

Biomedical Ultrasound/Bioresponse to Vibration: Tissue Scattering and Imaging

E. Carr Everbach, Chair
Swarthmore College, Dept. of Engineering, Swarthmore, PA 19081-1397

Contributed Papers

10:45

3aBBb1. Validating and expanding the theoretical frameworks that relate ultrasonic backscatter to scatterer properties. Shreedevi Dasgupta, Ernest Feleppa, Jonathan Mamou (Riverside Res. Inst., 156 William St., New York, NY 10038), Mark Rondeau and Harriet Lloyd (Cornell Univ., New York, NY)

Backscattered echo signals are related to scatterer properties in theoretical frameworks published independently by Lizzi and Insana. These frameworks are based on assumptions that may not consistently apply to tissue scatterers. Our study aims to confirm and expand the theory to include scattering under a variety of scatterer conditions. A 35-MHz transducer was used to scan suspended 12-μm-diam uveal melanoma cells. Using Fourier methods, spectral parameters were computed from ultrasonic echo data acquired from the suspension. Normalization was implemented in two ways. First, a glass-plate calibration spectrum was subtracted from the experimental data. Second, the spectrum acquired from suspended 20-μm beads was compared to the spectrum predicted by Faran's theory. The difference between expected and experimental spectra was attributed to system effects. Cell acoustic impedance was estimated from cell-pellet density and propagation velocity. Theoretical spectral values were computed from cell size, concentration, and acoustic impedance using the Lizzi framework. The cell pellet propagation velocity was 1550 m/s, and the acoustic impedance was 3587 kg/m-s. The normalized, experimental spectral parameters agreed closely with the theoretical predictions. Future work will include different scatterer sizes and geometries and a comparison of results obtained using the Lizzi and Insana frameworks.

11:00

3aBBb2. Histology-based simulation of ultrasonic scattering in cells and tissues. Timothy E. Doyle, Keith H. Warnick, and Brent L. Carruth (Dept. of Phys., Utah State Univ., 4415 Old Main Hill, Logan, UT 84322-4415, timdoyle@cc.usu.edu)

Previously developed computer programs for simulating ultrasonic scattering in cells and tissues at the microscopic level [T. E. Doyle and K. H. Warnick, J. Acoust. Soc. Am. **120**, 3283 (2006)] have been further refined and tested. The programs model the cells and nuclei in tissues as spherical particles with the use of vector multipole expansions and boundary condition solutions for the scattered fields. In addition, theorems, iteration, and matrix methods are used to solve for multiple scattering both inside and outside the cells. The latest refinements include the development of algorithms for simulating ultrasonic scattering from cells with nuclei arbitrarily located within the cell, and not centered as in previous models. Backscatter spectra have also been acquired for simulated cell clusters of up to several thousand cells. The effects of tissue structure, nucleus size, extracellular elastic properties, and intracellular elastic properties were investigated. The models provide a mechanistic link between measurable parameters in medical ultrasound and histological changes associated with various tissue pathologies. The application of these results to the improvement of ultrasonic tissue characterization and the specificity of various ultrasonic imaging modalities are discussed.

11:15

3aBBb3. Measurement of the anisotropy of the backscatter coefficient of formalin-fixed myocardial tissue. Min Yang, Todd Krueger, Mark Holland, and James Miller (Lab. for Ultrason., Dept. of Phys., Washington Univ., One Brookings Dr., Campus Box 1105, St. Louis, MO 63130)

The objective of this study was to measure the backscatter coefficient of myocardial tissue as a function of the angle of insonification relative to the predominant myocardial fiber direction. Eight cylindrical formalin-fixed ovine left-ventricular myocardial specimens were immersed in a water bath and insonified with a 5 MHz center frequency ultrasonic transducer. Backscattered ultrasonic data were acquired from the mid-myocardium of each tissue specimen over a full rotation. The backscattered data obtained for each angle of insonification were compensated for the effects of attenuation and diffraction to obtain measurements of the backscatter coefficient. Values for the backscatter coefficient at 5 MHz exhibited marked anisotropy, ranging from $(17 \pm 14) \times 10^{-4}$ cm^{-1} Sr^{-1} when insonified perpendicular to the predominant myofiber orientation to $(1.2 \pm 0.7) \times 10^{-4}$ cm^{-1} Sr^{-1} when parallel (mean $\pm$ standard deviation).

11:30

3aBBb4. Improving tissue-type imaging of prostate cancer by combining magnetic-resonance spectroscopy and ultrasound spectral parameters. Ernest Feleppa, Shreedevi Dasgupta, Sarayu Ramachandran, Andrew Kalisz (Riverside Res. Inst., 156 William St., New York, NY 10038), Christopher Porter, Marc Lacrampe, and Christina Isacson (Virginia Mason Medical Ctr., Seattle, WA)

This study aims to improve existing prostate tissue-type imaging methods by utilizing independent tissue properties sensed by MR and US. An artificial-neural-network classifier was trained using the blood PSA level and ultrasonic spectral parameters of echo signals derived from biopsied regions of 64 patients. Biopsy-core histology was used as the gold standard. Classifier performance was assessed using ROC analysis. We generated tissue-type images (TTIs) using a look-up table that returned cancer-likelihood scores for spectral parameter and PSA combinations. We assessed the feasibility of integrating MR and US parameters to improve classification. 3-D renderings of the prostate from transverse MR and US scans were generated from data acquired preoperatively. The 3-D rendering obtained from MR data showed that the prostate was considerably distorted (flattened) by the large MR endorectal probe compared to the distortion caused by the smaller US endorectal probe. The MR 3-D rendering was warped successfully to match the US volume. Warping successfully compensated for the deformations introduced during scanning. This suggests that accurate coregistration of 3-D data acquired using MR and US is possible. Such coregistration is essential for developing an improved classifier based on MR and US parameters.

11:45

3aBBb5. Classification of trabecular bone tissue through acoustic scattering. Max Denis, Charles Thompson, and Kavitha Chandra (Univ. of Massachusetts Lowell, One University Ave., Lowell, MA 01854)

In this work, the scattering of acoustic waves from a three-dimensional trabecular bone is considered. The trabecular bone volume is represented as a binary medium comprised of fluid and solid phases. A perturbation

expansion is performed in terms of gauge functions based on the density and compressibility contrast between the constituent components of the bone tissue. Internal resonance and multiple scattering effects are considered. In our simulations, trabecular bone samples having normal and osteoporosis condition will be examined. It is shown that the interior features of the media can be extracted by analysis of the forward scattered field. Of particular interest are features of anisotropy and frequency dependence of scattering field. The potential utility of this method for diagnosing osteoporosis stages in trabecular bone tissue is discussed.

WEDNESDAY MORNING, 6 JUNE 2007

CANYON ROOM A, 8:00 TO 10:45 A.M.

Session 3aNS

Noise: Aircraft Noise and General Topics in Noise

J. Micah Downing, Cochair
Blue Ridge Research and Consulting, 13 1/2 W. Walnut St., Asheville, NC 28801

Kent L. Gee, Cochair
Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602

Contributed Papers

8:00

3aNS1. Analysis of acoustical nonlinear measures for aircraft flyover data. Michael James, Micah Downing (Blue Ridge Res. and Consulting, LLC, 13 $\frac{1}{2}$ W. Walnut St., Asheville, NC 28801), Sally Anne McInerny (The Univ. of Alabama, Tuscaloosa, AL 35487), and Kent Gee (Brigham Young Univ., Provo, UT 84602)

Numerous analysis techniques have been proposed as means of characterizing acoustical nonlinearities in high-thrust engine noise. These include probability distributions for the pressure and the time derivative of the pressure (i.e., the gradient), the skewness and kurtosis coefficients of the pressure and its time derivative, bispectral analysis, and Howell-Morfey plots. In this presentation, a number of these analysis techniques are applied to acoustic data recorded from several different military jet flyovers. The results presented here emphasize the need for high bandwidth, large signal-to-noise ratio measurements.

8:15

3aNS2. Directivity and spectral characteristics of aircraft noise during landing operations. Bradley M. Dunkin, Anthony A. Atchley (Grad. Prog. in Acoust., The Pennsylvania State Univ., University Park, PA 16802), and Kathleen K. Hodgdon (The Pennsylvania State Univ., State College PA 16804)

Recordings of aircraft landing operations have been analyzed to obtain directivity and spectral characteristics of the noise emitted during these operations. The recordings were made with a widely spaced microphone array positioned near and parallel to the 19R runway at Washington-Dulles International Airport. In addition, the landings were recorded with a video camera to obtain positional data of the aircraft. Given an aircraft's position and velocity along the runway, the signal at each microphone is corrected for retarded time, spherical spreading, and Doppler shift. The corrected signals are combined to establish a directivity pattern, and third-octave analysis is performed to obtain spectral characteristics. The results of this analysis are intended to be used as input to airport noise modeling software such as the Integrated Noise Model (INM), which currently models landing noise using forward-thrust input data. The current analysis should provide more realistic directivity and spectra for the reverse-thrust condition. [Work supported by the Federal Aviation Administration and the Applied Research Laboratory Exploratory and Foundational Research Program.]

8:30

3aNS3. On the perception of crackle in noise radiated from military jet aircraft. Kent L. Gee (Dept. of Phys. and Astron., Brigham Young Univ., N319 ESC, Provo, UT 84602, kentgee@byu.edu), Victor W. Sparrow, Anthony A. Atchley, and Thomas B. Gabrielson (The Penn State Univ., University Park, PA 16802)

Crackle, a phenomenon sometimes associated with supersonic jet noise, has been usually quantified using the time waveform skewness. Skewness is a measure of the asymmetry of the waveform probability density function (PDF). This paper discusses whether skewness is a sufficient descriptor for the perception of crackle in jet noise waveforms. As part of this study, methods by which non-Gaussian waveforms may be generated have been investigated. By passing a Gaussian noise signal through an appropriately selected nonlinear transformation function, a simulated waveform can be created that has the same mean, variance, skewness, kurtosis, and power spectrum as a crackle-containing F/A-18E waveform recorded 18 m from the aircraft during a tie-down test. In addition, the PDFs of the simulated signal and the recorded signal are nearly identical. Despite the essential equivalence of the two waveforms from a statistical and spectral perspective, there is no perceived crackle when the simulated waveform is played back. Further analysis of the two waveforms suggests that perception of crackle is likely related to rapid changes in acoustic pressure that reveal themselves in the characteristics of the time derivative of the waveform than of the waveform itself. [Work supported in part by SERDP and ONR.]

8:45

3aNS4. Comparison of artificial neural network structures to identify military impulse noise. Brian A. Bucci and Jeffrey S. Vipperman (Univ. of Pittsburgh, 648 Benedum Hall, 3700 O'Hara St., Pittsburgh, PA 15261)

To monitor the production of military impulse noise in the area of military installation a classifier is being developed to identify military impulse events from other noise sources. In previous work [B. A. Bucci, J. S. Vipperman, J. Accoust. Soc. Am. **119**, 3384 (2006)], efforts were made to identify military impulse noise from other noise sources using a multi-layer perceptron neural network. The network used scalar input metrics (kurtosis, crest factor, spectral slope, and weighted square error) computed from recorded wave forms. In an extension to this effort, various types of neural network classifiers utilizing wavelet coefficients and cepstral coefficients as inputs are currently being tested and compared. These network types, which have proven successful in similar applications, include multi-layer perceptrons, radial basis function network, self-organizing maps, and recurrent networks. The networks are trained on an increased library of

waveforms as compared to the previous effort. The goal of this effort is to also produce a classifier continuous operation rather than analyzing selected recordings as in the previous effort. [This research was supported wholly by the U.S. Department of Defense, through the Strategic Environmental Research and Development Program (SERDP).]

9:00

3aNS5. Evolution of time-varying loudness during nonlinear propagation of broadband noise. S. Hales Swift and Kent L. Gee (Dept. of Phys. and Astron., Brigham Young Univ., N283 ESC, Provo, UT, 84602, hales.swift@gmail.com)

Previous work has shown that loudness and sound quality metrics that rely exclusively on the power spectrum do not adequately represent the perceptual impact of nonlinearly propagated noise. [K.L. Gee *et al.*, J. Acoust. Soc. Am. **121**, EL1–EL7 (2007)]. This discrepancy between spectral content and perceptual quality appears to be related to the formation of acoustic shocks in the pressure waveform. Because time-varying loudness (TVL) more accurately represents temporal effects in the ear, a preliminary investigation of the response of TVL to nonlinearly propagated noise has been carried out. TVL has been calculated for a shaped broadband noise waveform as it is nonlinearly propagated using a generalized Burgers equation-based model. The same waveform is also propagated using a linear model and the TVL for the nonlinear and linear waveforms are compared and discussed. An additional comparison carried out is that of the nonlinear waveform's TVL and the TVL for the nonlinear waveform with its Fourier phase randomized. This phase randomization preserves the nonlinear waveform's power spectrum but eliminates the acoustic shocks in the time domain. The nonlinear waveform shows more variation in both short and long term loudness than either the linear or the phase-randomized waveform.

9:15–9:30   Break

9:30

3aNS6. Measurement of vessel underwater noise, standard development status. Michael Bahtiarian (Noise Control Eng., 799 Middlesex Turnpike, Billerica, MA 01821, mikeb@noise-control.com)

The development of an entirely new commercial standard for underwater noise measurement of ships started in early 2007. Currently, no voluntary consensus standard exists for performing underwater noise measurements of ships. For many years, the field of underwater noise from ships has been the exclusive specialty of the Navy. However, non-Navy vessels are looking to be just as quiet so that they can perform better science. Green Ships are being conceived in order to have less emission into the ocean. The goal of the project is to develop an American National Standard for the measurement of underwater noise levels of ships using commercial technology. One aim is that the standard would be applicable to any open ocean site in the world and not require traveling to special acoustic test range. The committee's scope of work will include neither regulatory actions nor the development of any underwater noise level limits. This presentation will be an update of the committee work to date. A mission statement of the committee and discussion of measurement grades is to be presented. The presentation shall also serve as outreach to the acoustical community. As such, the author hopes to provide time for questions and feedback.

9:45

3aNS7. Modelling sound preference. Lei Yu and Jiang Kang (Architectural School, Univ. of Sheffield, Western Bank, Sheffield, S10 2TN UK, Lei.Yu@shef.ac.uk)

More and more attention is being paid to soundscape quality by today's acoustic researchers and environmental designers. It is important that, in addition to noise annoyance, sound preference is another main component to determine soundscape quality, especially in urban open public spaces. Previous studies showed that subject evaluation of sound preference in urban open public spaces varied according to physical, psycho-

logical, and social variables [W. Yang and J. Kang, J. Urban Des. **10**, 69–88 (2005)]. Prediction models using artificial neural network (ANN) techniques for sound preference have been developed, as presented in this paper. Suitable input variables for ANN models have been derived from SPSS analyses based on data collected from large scale field studies and laboratory experiments [L. Yu and J. Kang, J. Acoust. Soc. Am. **120**, 3238 (2006)]. These input variables include psychoacoustic indices such as loudness, sharpness and roughness, subject's social/demographic factors and their long-term sound experience, while the sound preference, described as acoustic comfort or pleasantness, is set as the output. Through the process of training and testing the models, good convergences have been achieved, which means that the prediction of sound preference evaluation is possible at the design stage.

10:00

3aNS8. Case study of soundscape assessment and design methods. Gary W. Siebein, Youngmin Kwon, Pattra Smitthakorn, and Martin A. Gold (Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611-5702)

A case study of the soundscape assessment and design process for an urban design intervention in a medium-sized city was conducted as an investigation of acoustical assessment and design methods. A series of long-term acoustical measurements of average sound levels in the community as well as more detailed measurements and aural recordings of specific acoustic events that comprised the ambient sounds were mapped for the community. Measurements included overall A-weighted and octave band sound levels and calibrated audio recordings. Sound walks were conducted at various times of day to understand the dynamics of the acoustical environment and to identify issues. Focus group discussions among stakeholders and team members developed the long-term plans for the community and determined appropriate architectural and acoustical design criteria for the project. The acoustical data were used as the basis for determining design strategies for the urban interventions including reducing, buffering, and mitigating undesirable existing sounds, preserving and enhancing desirable existing sounds, and designing new soundscape elements to enhance the comprehensive plan for the community. The results of the study included an assessment of the acoustical analysis methods and soundscape design strategies for the project.

10:15

3aNS9. Evaluation of field impact sound pressure levels as a function of tapping machine location. John J. LoVerde and D. Wayland Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Field impact sound insulation in the United States is typically quantified in terms of Field Impact Insulation Class, which is defined in ASTM E1007. This standard requires averaging the receiving room impact sound pressure levels (ISPL) generated by the tapping machine in four defined positions. To our knowledge the variation in ISPL with different tapping machine positions has not been quantified for field measurements. Hundreds of recent field tests are compiled for a statistically significant analysis of this variation, covering all common floor/ceiling assembly construction types and many different flooring and resilient materials. The effects of reducing the number of tapping machine positions in the ASTM standard on the accuracy and precision of the measurement are evaluated.

10:30

3aNS10. Global and local responses of coupled dynamic systems. Gideon Maidanik (NTI Corp., Bethesda, MD) and Ronald Hughes (NSWCCD, West Bethesda, MD 20817)

Responses of a master oscillator coupled to a set of satellite oscillators are evaluated. A set is designed by specifying a normalized frequency distribution. When the designed distribution is completely degenerate, the influence of the coupling is global. The coupling generates two global peaks, one on each side of an antiresonance that replaces the peak in the response of the uncoupled master oscillator. The loss factor for these global peaks matches that of the uncoupled master oscillator. As the degeneracy is gradually removed, the evaluated response gradually reveals,

within the distribution bandwidth, contributions from individual satellite oscillators. The loss factors and the frequency separation of the two global peaks then become dependent upon the coupling. As the degeneracy is further removed, the local presence of the satellite oscillators assumes the dominant feature in the response whereas the presence of the two global peaks subsides. Since the designed distributions relate loss factors directly to modal overlap parameters, increases in these parameters above unity render this dominance in the response quiescent and saturated. On the other hand, continuous decreases in these parameters below unity increase the fluctuations in that dominant part of the response *ad infinitum*.

WEDNESDAY MORNING, 6 JUNE 2007 ALPINE WEST, 8:00 TO 11:05 A.M.

Session 3aPAa

Physical Acoustics and Education in Acoustics: Physical Acoustics Demonstrations

James M. Sabatier, Cochair
Univ. of Mississippi, National Center for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Murray S. Korman, Cochair
U.S. Naval Academy, Physics Dept., 572 C Holloway Rd., Annapolis, MD 21402

Chair's Introduction—8:00

Invited Papers

8:05

3aPAa1. Faraday waves and ultrasonic foggers. R. Glynn Holt (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Locally flat liquid interfaces that are periodically forced to vibrate above a threshold amplitude exhibit an instability to waves of finite amplitude and half the forcing frequency. These waves (first observed and explored by Faraday) self-organize into a variety of patterns depending on experimental parameters. When driven at sufficient amplitude they can become turbulent and result in the pinch-off of liquid droplets. A very brief review of the relevant theory will be followed by a demonstration of acoustically induced Faraday waves in a couple of particular geometries. At least one demonstration of a practical application of Faraday waves will also be shown.

8:20

3aPAa2. Demonstration of a highly directional acoustic source. Preston S. Wilson, Wayne M. Wright, and Mark F. Hamilton (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Highly directional light sources such as flashlights, spot lights, and lasers are common objects well known to most people. In contrast, a highly directional acoustic source, or, in other words, a source of sound that is audible in only a very narrow region of space, is far less common. Most people have never experienced such a source, and the phenomenon is not found in nature. A highly directional source of sound known as a parametric array is used underwater for sonar applications, but the frequency (pitch) of the sound is often above the hearing range for human listeners. Similarly, highly directional, focused sound sources are regularly used in medical applications, but, again, the frequency is too high to be heard. The narrowness of the acoustic beam cannot be experienced by human listeners. Recently, parametric array technology has been commercialized for use in air at frequencies in the human auditory range. These devices produce very narrow beams of audible sound. When pointed at one person, who can be listening contentedly, the sound is inaudible to another person only a few feet away. Such a device will be demonstrated and the basic physics behind its operation will be explained.

8:35

3aPAa3. Demonstration of nonlinear acoustic landmine detection: Hysteresis effects observed in a soil-elastic plate oscillator. Murray S. Korman, Kathleen E. Pauls, Sean A. Genis (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402), and James M. Sabatier (Univ. of Mississippi, University, MS 38677)

In nonlinear acoustic landmine detection the vibration interaction of the top-plate of the buried plastic landmine with the adjacent soil is a subject of interest due to the strong nonlinear coupling. Donskoy discovered the significance of using nonlinear vibro-acoustic two-tone tests to help find buried relatively compliant landmines at certain nonlinearly generated combination frequencies; and that false alarms were reduced due to minimal nonlinear interactions over relatively rigid buried objects. Later, work by Sabatier showed that certain nonlinear experimental results involving tuning curves and profiling at a large number of combination frequencies from two-tone tests (in buried plastic anti-tank inert landmine experiments) revealed that the nonlinear effects were similar to the mesoscopic/nanoscale nonlinear elastic behavior observed in geomaterials like sandstone. Here, an apparatus called the soil-plate oscillator is used to model the nonlinear interaction of a thin clamped vibrating acrylic plate with a small column of dry sifted soil supported over it. Measurements of the effects of tuning curves taken at increased amplitude acoustic drive amplitude (exhibiting softening) can be explained if hysteresis effects (characteristic of the mesoscopic nonlinear) interaction are taken into consideration. Lissajou patterns of the driving force vs. displacement also reveal hysteresis effects. [Work support by ONR.]

8:50

3aPAa4. Visualizing acoustic resonances in quasi 2D solids. J. R. Gladden (Dept. of Phys. and Astron., Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

Acoustic resonances in solid objects are easily detected through the sound that they radiate, but the vibration patterns are hard to visualize because the frequencies are high and amplitudes are low. I plan to demonstrate a system using a birefringent solid which is acoustically driven at resonances frequencies. The birefringence allows strain fields to be visualized as bright and dark bands when the resonator is placed between crossed optical polarizing plates and illuminated from behind. Using a strobe light as the light source with the frequency closely matched to the resonance frequency slows the time scale of the acoustic cycle so it can be easily followed by the human eye.

9:05

3aPAa5. Ultrasonic signatures of human motion. Alexander Ekimov and James M. Sabatier (Univ. of Mississippi, NCPA, 1 Coliseum Dr., University, MS 38677, aekimov@olemiss.edu)

Human footsteps generate periodic broadband frequency vibrations in the ground/floor and sound in the air from a few hertz up to ultrasonic frequencies due to repeatable cycles of surface loading by dynamic forces. The friction force of a footstep on the ground/floor produces broadband ultrasonic frequencies. Human motion can be characterized as a periodic, temporal process of a multi-degree of freedom mechanical system. The individual human body parts (torso, legs, arms, etc.) have different cross-sections and Doppler signatures that can be used for human recognition among other moving and stationary objects. A demonstration of human motion measurements using passive and active ultrasonic methods will be presented. The passive method employs ultrasonic measurements by a narrow band microphone that is sensitive to the sound from sliding contacts. The active method utilizes continuous wave Doppler ultrasound. Dynamic signatures of a walking person by the two methods will be demonstrated in real time during the presentation. [Work supported by Department of the Army, Army Research Office contract W911NF-04-1-0190.]

9:20

3aPAa6. Angular dependence of target strength for simple geometric shapes. Edward Tucholski (Phys. Dept., Chauvenet Hall, U.S. Naval Acad., 572C Holloway Rd., Annapolis, MD 21402)

Understanding the active sonar equation is a fundamental skill for a course in underwater acoustics and sonar such as that taught at the United States Naval Academy. Students must develop an intuitive understanding of each term in both the noise-limited and reverberation-limited cases. In this demonstration, the angular dependence of the target strength term is explored using commercially available 25- and 40-kHz airborne piezoelectric transducer/receiver pairs. By rotating standard target shapes (e.g., rectangular plates, circular plates, finite cylinders, ellipsoids), reflected intensity is measured as a function of angle of rotation. The measured data are compared with those predicted by standard theoretical predictions. While the same measurements could be made in a tank using submerged transducers, the airborne transducers offer the advantage of lower cost and simplicity of apparatus. Additionally, unwanted modal interference from sound reflection off tank surfaces is reduced in the airborne measurement. By direct observation, future naval officers are able to study target strength so as to understand how to position their sensors to optimize their own active sonar performance and position their warship to minimize detection by enemy active sonar.

Contributed Papers

9:35

3aPAa7. Demonstration of properties of human hearing using a cochlear analogue. Seth S. Tomlinson, Ryan Anderson, Kent L. Gee, and Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., N283 ESC, Provo, UT 84602, tom30000@byu.edu)

A device has been constructed that demonstrates the spectral decomposition of waves according to spatial location, analogous to the function of the basilar membrane of the mammalian cochlea. The design is based on a previous device constructed by Keolian [J. Acoust. Soc. Am. **101**, 1199–1201 (1997)]. We have modified Keolian's design to incorporate variable mass as well as stiffness and an electronically driven shaker, which allows the device to be driven with complex waveforms. In addition to showing spectral decomposition, the model can be used to demonstrate beating, masking, and other phenomena that occur on the basilar membrane. A demonstration of the device will be given.

9:50

3aPAa8. Live demonstration of a physical model of the active cochlea. Tina Jovic, Curtis King, and Richard Rabbitt (Dept. of Bioengineering, Univ. of Utah, 20 South 2030 East, Salt Lake City, UT 84112)

An active physical model of the cochlea was constructed to mimic the nonlinear amplification carried out by the human cochlea. The model includes hair-cell cilia-like sensors to measure the local velocity of the fluid, and electromagnetic actuators to actively feed energy back into the motion. The model is outfitted with an electromagnetic shaker (mimicking the eardrum) to deliver complex waveforms to a piston (mimicking the stapes). Since the model is much larger than the cochlea, its best response frequencies are shifted down in the spectrum. The most interesting responses are obtained when music is recorded in stereo, digitally processed, and replayed into the model. One channel is frequency shifted to match the best frequencies of the model while the other channel is played acoustically without modification. This allows the listener to observe the motion of the cochlear partition in temporal synchrony with hearing the original soundtrack. A live demonstration of the apparatus in response to pure tones, speech, and music will be presented under both passive and active conditions. [Work supported in part by NIDCD R01 DC04928.]

10:05−11:05
Discussions and Demonstrations

Session 3aPAb

Physical Acoustics and Engineering Acoustics: Acoustic Probes of Planetary Environments

Andi G. Petculescu, Chair

Univ. of Louisiana at Lafayette, Dept. of Physics, P. O. Box 44210, Lafayette, LA 70504

Chair's Introduction—9:00

Invited Papers

9:05

3aPAb1. Planetary exploration using acoustic instrumentation: The past and the future. Martin C. Towner (PSSRI, Open Univ., Walton Hall, Milton Keynes, MK2 2UA, UK, martintowner@googlemail.com)

To date only a couple of planetary missions have deployed acoustic instruments, primarily intended to probe acoustic environments (such as listening for thunder) rather than to investigate physical properties. The recent Huygens probe mission to Saturn's moon Titan is the first to carry our more involved measurements, but still using relatively simple instruments as an altimeter and to estimate the speed of sound. I briefly review some of the acoustic instruments that have been flown on previous space missions, and the environments they have had to deal with, and discuss in more detail the Huygens results. Spacecraft designers are by nature conservative, tending to use older, more reliable technology, but capable instruments are still possible, and it is hoped to show how small simple acoustic sensors can contribute a lot to the scientific understanding for relatively few resources. Forthcoming opportunities are also considered, based on current exploration plans.

9:30

3aPAb2. Development of the first audio microphone for use on the surface of Mars. Gregory T. Delory, Janet Luhmann (Space Sci. Lab., Univ. of California, Berkeley, CA 94720), Louis Friedman, and Bruce Betts (The Planetary Society, Pasadena, CA 91106)

In 1999, NASA launched the Mars Polar Lander (MPL) to study volatiles and the climate history of Mars. Among the many instruments aboard the spacecraft was the first audio microphone sensor ever developed for the environment of Mars, included as a piggy-back sensor on the Russian LIDAR experiment. Funded by The Planetary Society and built and tested by the U.C. Berkeley Space Sciences Laboratory, the Mars Microphone was a unique and powerful tool for public engagement with the Mars exploration program. The possibilities for sound sources on Mars included wind and dust found in dust devils and storms, atmospheric electricity, and lander-generated sounds such as camera movements and the digging of the robotic arm. Here we outline the design, development, and test program for a spaceflight acoustic sensor, where the unique challenges of extreme environments, limited data bandwidth, and requirements for a low-power, light-weight system were overcome. While the MPL mission was lost and no data ever returned, the Mars Microphone nonetheless stimulated interest in the science of extra-terrestrial acoustics and in future planetary mission instruments based on acoustic methods.

9:55

3aPAb3. The propagation of sound in planetary atmospheres and its application to bolide detection. Jean-Pierre Williams and Ian McEwan (Div. of Geological and Planetary Sci., California Inst. of Technol., 1200 E. California Blvd., Pasadena, CA 91125)

The behavior of sound (i.e., its radiation pattern and range of detectability) is characterized in the complex dynamical systems of the Venus, Earth, Mars, and Titan atmospheres as modeled by a global circulation model (GCM). Sound propagation is affected by the composition, structure, and dynamics of a planet's atmosphere; therefore, a modular approach is adopted so that any hypothetical atmosphere with varying climatic conditions and gas composition can be accounted for. Using the GCM to determine the structure and dynamics of an atmosphere, the sound path is ray traced accounting for viscous, thermal, and molecular attenuation. As a specific application, we discuss the practicality of an *in situ* acoustic sensor as a method for direct measurement of bolide influx rates in planetary atmospheres. Bolides provide well-characterized, large magnitude, low-frequency sound sources that are common to all planetary atmospheres and should be remotely detectable by acoustic sensors. End-to-end comparisons between Venus, Earth, Mars, and Titan are made by modeling a typical large terrestrial event and the resulting sound propagation in each environment. [Work funded by the O.K. Earl postdoctoral fellowship.]

10:20−10:35   Break

10:35

3aPAb4. The Huygens probe Surface Science Package acoustic instrument: Results at Titan. John C. Zarnecki (Planetary & Space Sci. Res. Inst., The Open Univ., Walton Hall, Milton Keynes, MK7 6AA, UK)

The European Space Agency's Huygens probe, when it landed on the surface of Titan, Saturn's largest moon, on January 14, 2005, carried the first active acoustic sensors to be deployed on another cosmic body. The Surface Science Package contained a simple sonar (operating at 14.7 kHz) and a speed of sound sensor (operating at 1 MHz). The design and operation of the sensors will be described.

Despite being optimized for the case of landing on a surface of liquid hydrocarbons, the landing on a solid surface still produced valuable data. The former sensor obtained echoes from Titan's surface during the last 90 m of the descent. Interpretation of these data will be discussed in terms of surface properties, such as terrain topography and possible physical nature, around the landing site. Probe landing speed was also derived with high precision. The speed of sound instrument operated in the lower 11 km of the atmosphere. These data have been used to constrain the methane content of Titan's atmosphere to approximately 2% at 10 km rising to about 3.5% at lower altitudes.

Contributed Papers

11:00

3aPAb5. Alien soundscapes: Acoustics on Titan, Venus, Mars, and Earth. Andi Petculescu (Dept. of Phys., Univ. of Louisiana at Lafayette, Lafayette, LA 70504, andi@louisiana.edu) and Richard Lueptow (Northwestern Univ., Evanston, IL 60208)

In January 2004, the general public was able to listen to the sounds of an alien world for the first time. The world was Saturn's moon, Titan, and the sounds were audio recordings taken during the descent of the Huygens probe. In this paper, calculations of sound speed and attenuation in the atmospheres of Titan, Venus, Mars, and Earth are obtained using a new molecular relaxation-based model. Vertical atmospheric profiles of sound speed and attenuation are calculated at a frequency of 15 kHz. The profiles are obtained using currently available data of temperature, pressure, and density. The sound speed profiles mirror the differences in atmospheric layering. The absorption profiles are comparatively smooth for Mars, Titan, and Earth while Venus has a noticeable attenuation dip between 40 and 100 km. During descent of a probe, acoustic quiet zones occur when sensing sound is produced by a low-frequency simple source near the surface. The frequency dependence of sound speed and attenuation is calculated at the surfaces of the four bodies. The effect of acoustic atmospheric filtering on an audio sample transmitted in the four planetary environments is also addressed. [Work funded by NASA.]

11:15

3aPAb6. The propagation of sound on Titan using the direct simulation Monte Carlo. Amanda D. Hanford, Lyle N. Long, and Victor W. Sparrow (Grad. Program in Acoust., The Penn State Univ., University Park, PA 16802)

With the recent success of the Huygens lander on Titan, a moon of Saturn, there has been renewed interest in exploring the acoustic environment of the only moon in the solar system with a significant atmosphere. The direct simulation Monte Carlo (DSMC) method is used here for mod-eling sound propagation in the Titan atmosphere. DSMC is a particle method that describes gas dynamics through direct physical modeling of particle motions and collisions. DSMC is based on the kinetic theory of gas dynamics, where representative particles are followed as they move and collide with other particles. The validity of DSMC for the entire range of Knudsen numbers (Kn), where Kn is defined as the mean free path divided by the wavelength, allows for the exploration of sound propagation in the Titan atmosphere for all values of Kn. DSMC results for the absorption of sound have shown that sound absorption depends heavily on Kn and deviates significantly from the continuum classical assumption for large Kn. This talk presents the successful application of the DSMC method to acoustic waves in the Titan atmosphere for a large range of Kn. [Work supported by NASA.]

11:30

3aPAb7. A small logging sensor station for infrasonic detection of bolides to complement a camera network. Martin C. Towner (PSSRI, Open Univ., Walton Hall, Milton Keynes, MK7 6AA, UK, martintowner@gmail.com) and Philip A. Bland (Imperial College, London, SW7 2AZ, UK)

A small simple logging station is described that is intended to search for and detect the infrasonic signature of an incoming meteor. This is part of activities to support a fireball camera network currently being deployed in the Nullarbor desert in Western Australia. Camera networks are constrained in the time that they can be active by their required viewing conditions; for example they are not generally active during daylight hours or when the weather is overcast. As a complementary approach, investigation has been carried out concerning the detection of meteors by alternative means, most notably by investigating their infrasonic and electromagnetic signature. This system consists of a small logger, an infrasonic detector, a VLF detector, and other supplementary sensors, independently powered by solar panels, and with a satellite modem for data return.

Session 3aPP

Psychological and Physiological Acoustics: Binaural Hearing and Processing of Complex Sounds

Beverly A. Wright, Chair

Northwestern Univ., Dept. Communication Science and Disorders, 2240 Campus Dr., Evanston, IL 60208

Chair's Introduction—8:00

Contributed Papers

8:05

3aPP1. Naive performance on interaural time difference discrimination with high-frequency amplitude-modulated tones. Yuxuan Zhang and Beverly A. Wright (Dept. of Commun. Sci. and Disord., Inst. for Neurosci., Northwestern Univ., Evanston, IL 60208, y-zhang6@northwestern.edu)

Humans are sensitive to changes in interaural time differences (ITDs) in high-frequency sounds with fluctuating amplitude envelopes. However, despite noted large individual differences in high-frequency ITD discrimination ability, to date, most investigations of this performance have employed only a few, typically experienced, listeners. Thus, it is not clear whether the data thus obtained are representative of the general, naive, population. Here, ITD discrimination thresholds with high-frequency amplitude-modulated tones were examined in nineteen listeners, all naive to psychoacoustic testing. The large range of performance across these listeners accommodated the collective data previously reported, but was not fully illustrated by any single previous investigation with a small sample size. For these naive listeners, ITD discrimination performance at high frequencies, 1) was much worse than that at low frequencies, 2) was better with a slower (0.15 kHz) than a faster (0.3 kHz) modulation rate with a 4-kHz carrier, and 3) did not differ between a lower (4 kHz) and higher (6 kHz) carrier frequency with a 0.3-kHz modulation rate. These data help establish high-frequency ITD discrimination performance in the general population, and may aid in the interpretation of performance measured in situations where testing time is limited. [Work supported by NIH.]

8:20

3aPP2. Within- and across-listener performance variability in discrimination of interaural time and level differences. Yuxuan Zhang and Beverly A. Wright (Dept. of Commun. Sci. and Disord., Inst. for Neurosci., Northwestern Univ., Evanston, IL 60208, y-zhang6@northwestern.edu)

The ability to discriminate changes in interaural time differences (ITDs) and interaural level differences (ILDs) varies markedly both within and across individuals. Such variability has been frequently noted, but rarely systematically investigated. Here, both forms of variability, expressed in terms of coefficients of variation (mean divided by standard deviation), were examined across a variety of ITD and ILD conditions in a total of 84 naive listeners. Three consistent patterns emerged. First, across-listener variability was relatively constant across ITD and ILD discrimination, suggesting that this variability might be determined by factors common to the processing of both cues. Second, compared to across-listener variability, within-listener variability was larger for the ITD conditions but smaller for the ILD conditions, suggesting that, within individuals, ITD processing may have a higher level of internal noise than ILD processing. Third, stimulus frequency (ranging from 0.3 to 6 kHz) and stimulus type (pure tones and sinusoidally amplitude modulated tones) had little influence on either type of variability, suggesting that variability was determined by cue-dependent, rather than stimulus-dependent mechanisms. The systematic nature of these variability patterns illustrates that examination of variability can provide important insights into the processing of interaural cues. [Work supported by NIH.]

8:35

3aPP3. Modeling the detection of interaural incoherence. Matthew J. Goupell and William M. Hartmann (Phys. and Astron., Michigan State Univ., East Lansing, MI 48824)

Several hundred reproducible noises, with interaural coherence slightly less than 1.0, were used in psychoacoustical experiments intended to discover how listeners detect small amounts of interaural incoherence in narrow-band noise stimuli. The most successful detection model, which correlated with the detection data at $r = 0.87$, employed independent encoding of compressed (exponentially saturating) interaural phase fluctuations (IPF) and compressed interaural level fluctuations (ILF). Fluctuations were recombined centrally such that 5 degrees of IPF (28 μs of ITD) were equivalent to 1 dB of ILF. Integration times were found to be less than 10 ms, indicating that binaural sluggishness played no role. Neurophysiologically motivated short-term cross-correlation models (including auditory filtering, parameterized cochlear compression of the envelope, half-wave rectification, and temporal integration) were somewhat less successful in that they led to r values that were never greater than 0.78. Models were tested in step-by-step calculations to quantify the consequences of the various components of the models. [Work supported by the NIDCD of the NIH, Grant DC-00181.]

8:50

3aPP4. A stable capacity limit in interaural time-difference discrimination across training and individual listeners.. Beverly A. Wright and Yuxuan Zhang (Dept. of Commun. Sci. and Disord. and Inst. for Neurosci., 2240 Campus Dr., Northwestern Univ., Evanston, IL 60208-3550, b-wright@northwestern.edu)

An interesting learning pattern was previously observed for ITD-discrimination training with a 500-Hz tone. While the highest of the twelve threshold estimates obtained each day from each listener (n=8) decreased significantly over days, the lowest did not. However, the lowest values averaged ~6 microseconds, raising the question of whether they represented actual discrimination ability or resulted from random walks in the adaptive testing procedure. In an attempt to address this question, all twelve estimates of each listener were ordered from lowest to highest and the effect of training was examined separately for each rank. There was no improvement for the four lowest ranks, which ranged from ~6–25 microseconds. These values represent 33% of the data, a seemingly unlikely rate for random walks. In addition, in naive listeners (n=32), despite marked inter-individual variation in the average and highest thresholds, the lowest thresholds were constant across listeners. Though these analyses do not preclude the potential contribution of random walks within this data set, they keep open the possibility that the invariance in the lowest thresholds through training and across listeners was not an artifact of the adaptive procedure, but rather reflects a stable capacity limit in ITD processing. [Work supported by NIH.]

3118   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   3118

3aPP5. Age-related declines in the precedence effect and the detection of the dynamic change in interaural correlation. Juan Huang, Ying Huang, Xihong Wu, and Liang Li (Dept. of Psych., Speech and Hearing Res. Ctr., Natl. Lab. on Machine Percept., Peking Univ., Beijing 100871 China, j.huang@pku.edu.cn)

Elderly adults experience greater difficulties with sound localization tasks involving the precedence effect (PE) than young adults (Cranford *et al.*, 1990). In this study, the echo threshold of the PE and the threshold for detecting the dynamic change in interaural correlation were investigated in 12 young adults and 8 elderly adults with clinically normal hearing. For studying the threshold of detecting the dynamic change in interaural correlation, participants were presented with either binaurally-uncorrelated or binaurally-anti-phase noise fragments, which were inserted in the middle of two interaurally identical (correlated) broadband noise markers (duration=900 ms) delivered by headphones. Either the duration of the noise fragment or the ITD of the noise markers was manipulated. Compared to young participants, elderly participants had both higher echo thresholds and higher fragment-duration thresholds at various ITDs. The echo threshold was significantly correlated with the fragment-duration threshold in elderly participants but not in young participants. These results suggest that there are age-related declines in processing interaural signals, which may account for elderly adults' poor performance in discriminating and localizing sound sources in noisy, reverberant environments.

9:20

3aPP6. Selective spatial attention in a dynamic cocktail party task: Evidence for a strategy based on masker minimization. Douglas Brungart, Nandini Iyer, and Brian Simpson (Air Force Res. Lab., 2610 7th St., WPAFB, OH 45433, douglas.brungart@wpfab.af.mil)

In cocktail-party listening tasks with multiple spatially separated talkers, it is not clear whether the performance advantages of spatial separation are the result of a target-max strategy, where the listener uses selective attention to enhance sensitivity at the location of the target, or a masker-min strategy, where the listener selectively suppresses sensitivity at the locations of the maskers. In this experiment, listeners were asked to attend to a randomly moving CRM target phrase that was masked by two independently moving CRM masking phrases. The competing signals were lateralized solely on the basis of ITD, and they were band-pass filtered into non-overlapping bands to minimize the effects of energetic masking. The results of the experiments showed (1) that listeners were much more sensitive to changes in the masker locations than to changes in the target location; (2) that listeners consistently performed worse when the two maskers were spatially separated than when they were co-located, even when the average separation between the target and the maskers increased; and (3) that spatial separation of the second masker provided almost no benefit when the first masker was co-located with the target speech. These results seem to provide general support for the masker-min theory of auditory spatial attention.

9:35–9:50 Break

9:50

3aPP7. Using auditory saliency to interpret complex auditory scenes. Varinthira Duangudom and David V. Anderson (Dept. of Elec. and Computer Eng., Georgia Inst. of Technol., Atlanta, GA 30332)

In this paper, a computational model for predicting preattentive, bottom-up auditory saliency is proposed. The model determines perceptually what in a scene stands out to observers. The vision equivalency of this is visual saliency as defined by Koch and others [Itti *et al.*, IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259, (1998)]. The proposed auditory saliency model is based on suppression or promotion of different feature maps obtained using auditory spectro-temporal receptive fields. The auditory saliency model does well in predicting what is salient for some common auditory examples and produces results that match well

with results obtained from psychoacoustic experiments. Additionally, experiments performed to evaluate the model found that there is a strong correlation between auditory scenes chosen as salient by the model and scenes that human subjects selected as salient. [Work supported by NSF.]

10:05

3aPP8. A perceptual space that can explain the robustness of bio-acoustic communication. Roy D. Patterson, Ralph van Dinther (Ctr. for the Neural Basis of Hearing, Dept. of Physiol., Development, and Neurosci., Univ. of Cambridge, Cambridge, UK, rdp1@cam.ac.uk), and Toshio Irino (Wakayama Univ., Wakayama, Japan)

The sounds that animals use to communicate, including the syllables of speech, have a very special "pulse resonance" form, which automatically distinguishes them from background noise. The parts of the body used to produce these sounds grow as the animal grows. Thus, there is "acoustic scale" variability in communication sounds, which poses a serious problem for the perception and recognition stages of communication. The success of bio-acoustic communication suggests that the auditory system has a special preprocessor that automatically normalizes for acoustic scale as it constructs our internal "auditory image" of a sound. In this paper, we propose that the normalization is performed in the early stages of the auditory pathway by adding an extra, rather special, dimension to the space of auditory perception. This paper is about the mathematics of the space, which has to be "scale-shift covariant" to support communication, and the discovery of a unitary operator that can construct the appropriate space. The mathematics makes it clear that there is no equivalent means of scale normalization available in the traditional time-frequency space of the spectrogram. [Research supported by the UK MRC (G0500221, G990369).]

10:20

3aPP9. Comparison of relative and absolute judgments of speaker size based on vowel sounds. Thomas C. Walters, Phil A. Gomersall, Richard E. Turner [a], and Roy D. Patterson (Ctr. for the Neural Basis of Hearing, PDN, Univ. of Cambridge, Downing St., Cambridge, CB2 3EG, UK, tcw24@cam.ac.uk)

Judgments of speaker size are largely determined by two acoustic variables: glottal pulse rate (GPR) and resonance scale [i.e., vocal tract length (VTL)]. Both variables change with age (or height), but the rate is governed by different factors. The interaction of the variables was previously measured using absolute judgments of speaker size [D.R.R Smith and R.D. Patterson, J. Acoust. Soc. Am. **118**, 3177–3186 (2005)]. The resulting size surface (over the GPR-VTL plane) bends down outside the normal range. In this paper, a method is developed for deriving the surface using size-discrimination data. In a two-alternative forced-choice experiment, listeners compared sequences of vowels scaled in GPR and VTL to represent speakers with slightly different sizes; they were required to choose the interval with the smaller speaker. Comparisons about a point in the plane reveal the gradient vector, and the vectors across the GPR-VTL plane can be integrated to estimate the size surface. The results indicate that the size surface would be essentially planar if determined by size discrimination. This indicates that relative size judgments are different from absolute size judgments, probably because some source knowledge is required for the absolute judgements. [Research supported by the UK Medical Research Council (G9900369).] [a] Currently at the Gatsby Computational Neuroscience Unit, Alexandra House, 17 Queen Square, London, WC1N 3AR, UK.

10:35

3aPP10. Implementing synthesis control using timbral adjectives. Alastair C. Disley, David M. Howard, and Andrew D. Hunt (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK)

The authors have previously developed theories of relationships between timbral adjectives and spectral features, with a goal of developing a synthesizer using timbral adjectives as synthesis controls [Disley *et al.*, J. Acoust. Soc. Am. **119**, 3333 (2006) abstract only]. These theories have

been applied to a software-based additive synthesizer developed in the freeware Pure Data environment, in conjunction with practicing musicians and music technologists. Realistic yet highly configurable synthesis as a foundation for timbral manipulation is permitted by the use of harmonic envelopes derived from successive Fourier analysis of real-world samples. Theories of timbral adjectives' relationships to spectral and time-domain features can be mapped onto synthesis primitives. User feedback can then be used in an iterative process to refine existing theories and develop new controls unique to an individual's personal understanding of one or more timbral adjectives. The qualities of the resulting synthesis system and some revised theories of timbral adjectives are both discussed.

10:50

3aPP11. Intervention for restricted dynamic range and reduced sound tolerance. Monica Hawley, LaGuinn Sherlock, Susan Gold, Allyson Segar, Christine Gmitter, Justine Cannavo, and Craig Formby (Dept. of Otorhinolaryngolo.—HNS, Univ. of Maryland Tinnitus and Hyperacusis Ctr., 16 S. Eutaw St., Baltimore, MD 21201)

Hyperacusis is the intolerance to sound levels that normally are judged acceptable to others. The presence of hyperacusis (diagnosed or undiagnosed) can be an important reason that some persons reject their hearing aids. Tinnitus retraining therapy (TRT), originally proposed for the treatment of persons with debilitating tinnitus, offers the significant secondary benefit of increased loudness discomfort levels (LDLs) in many persons. TRT involves both counseling and the daily exposure to soft sound from bilateral noise generator devices (NGs). We implemented a randomized, double-blind, placebo-controlled clinical trial to assess the efficacy of TRT as an intervention for reduced sound tolerance in hearing-aid eligible persons with hyperacusis and/or restricted dynamic ranges. Subjects were assigned to one of four treatment groups: (1) full treatment, both counseling and NGs, (2) counseling with placebo NGs, (3) NGs without counsel-

ing, and (4) placebo NGs without counseling. They were evaluated at least monthly, typically for 5 months or more, on a variety of audiometric tests, including LDLs, the contour test for loudness for tones and speech, word recognition measured at each session's comfortable and loud levels, and on electrophysiological measures. We will present interim results and selected examples of positive treatment effects. [Work supported by NIH R01 DC04678.]

11:05

3aPP12. The "noise" of tinnitus and the vagaries of perception. Phyllis Gildston (Brooklyn College of the City Univ. of New York, North Shore Ctr., 275 Middle Neck Rd., Ste. B-6, Great Neck, NY 11023)

Multiple suspected etiologies for tinnitus encourage the development of multiple rehabilitation approaches. Loud noises, drugs, hearing loss, ear infections, tumors, Meniere's, head injuries, and TMJ have been cited, among many others, as potential causes. Even high impact exercise and stress are suggested culprits. A full understanding of the phenomenon is still, however, a long way off. Once all known remediable sources have been attended to, why should a remaining noise in the ear (measured as perceived at a particular loudness level) cause such severe psychological consternation in so many sufferers while others barely notice it exists? What mechanisms may be involved in the alteration of perception following successful rehabilitative procedures such as masking, hearing amplification, cognitive behavioral perceptual retraining, or hypnotherapy? This presentation will focus on changes induced by hypnosis which typically lead to a significant decrease in the perceived loudness level of the tinnitus and a significant increase in psychological well being. What is known of the science and art of hypnosis will be explored in an attempt to explain this often achieved metamorphosis in perception. Video clips from hypnotherapeutic sessions with correlated perceptual aftermaths will complement a review of scientifically known facts about the disorder.

WEDNESDAY MORNING, 6 JUNE 2007 ALPINE EAST, 8:00 TO 11:10 A.M.

Session 3aSC

Speech Communication: Computational and Experimental Approaches to Fluid Dynamics of Speech Production

Anders Lofqvist, Cochair
Lasarettet, Dept. of Logopedics and Phoniatrics, S-221 85 Lund, Sweden

Christine H. Shadle, Cochair
Haskins Laboratories, 300 George St., New Haven, CT 06511

Chair's Introduction—8:00

Invited Papers

8:05

3aSC1. Dynamics of glottal airflow. Michael Krane (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804)

The traditional assumptions concerning airflow in the glottis during vocal fold vibration are assessed. An equation of motion describing the dynamics of glottal airflow, and which explicitly includes terms due to separation point motion, wall displacement, skin friction, and glottal jet inertia, is derived. The relative order of magnitude of each of these terms is assessed using experimental data. It is shown that there are two temporal regimes during each cycle of vibration. The first holds when the glottis is just opening and just closing, in which wall motion and friction are important. In the second regime, when the glottis is wide open, the traditional assumptions hold as long as glottal jet inertia is negligible. If jet inertia is dynamically relevant, the aerodynamics is unsteady. If not, then effects such as the Coanda effect are dynamically insignificant. Evidence for the dynamic relevance of glottal jet inertia is presented. [Authors gratefully acknowledge support from NIH Grant 2 R01 DC005642-04A1.]

3aSC2. Physical mechanisms of phonation onset: The role of flow instabilities. Zhaoyan Zhang, Juergen Neubauer, and David Berry (UCLA School of Medicine, 31-24 Rehab Ctr., 1000 Veteran Ave., Los Angeles, CA 90095)

Current models of phonation exhibit a series of flow phenomena such as an oscillating flow separation point, the Coanda effect, and vortex shedding downstream of the glottis. While accounting for such phenomena significantly increases the computational load and may influence many aspects of phonation, the influence of such airflow features on phonation onset is still unknown. In a two-dimensional continuum model of phonation with a fixed flow separation point, a recent analysis showed that, in the absence of acoustics, the primary physical mechanism of phonation onset was the matching of the flow stiffness (i.e., the intraglottal airflow pressure) with the elastic stiffness of the vocal folds, resembling common observations of human phonation. Thus, the oscillation of the flow separation point may be merely a consequence of the unsteady motion of the vocal folds, with only a minor influence on phonation onset at low frequencies. With a natural frequency much higher than the phonation frequency, flow instabilities such as vortex shedding may also only play a secondary role in determining the conditions of phonation onset (apart from being responsible for flow separation and the transglottal pressure drop), at least for normal phonation. Further experiments are needed to clarify the relative importance of these flow features on phonation onset.

3aSC3. Fluid dynamics measurements during phonation. Fariborz Alipour (Univ. of Iowa, Dept. Sp. Pathol. and Audiol. Iowa City, IA 52242)

Fluid dynamic properties of air within the larynx changes so rapidly during phonation that its measurement could be a challenge for voice research. These measurements included time varying glottal pressure and air particle velocity in narrow passages that pose many technical difficulties. In this paper, to elucidate these challenges, measurement of dynamic pressure acting on the vocal folds in a canine hemilarynx model has been initially examined. Then static and dynamic volume flow measurements and, finally, the particle velocity measurements with hot-wire anemometry within the glottal jet have been discussed. Canine excised larynges as well as canine hemilarynges were used as phonation models. For dynamic pressure measurements, miniature pressure transducers were embedded in a Plexiglas plate that replaced one vocal fold and recorded the oscillation pressure. All mounted laryngeal models were subjected to increasing pressure-flow sweeps at various adduction levels and the measured data was processed for dynamic pressure and velocity profiles. Results indicated that pressure within the glottis vary both vertically and longitudinally during each vibratory cycle. The pressure signals close to the superior edge showed large negative values due to the motion of the vocal folds. [Work supported by NIDCD Grant No. DC03566.]

3aSC4. Intraglottal pressures relative to laryngeal airway configuration. Ronald C. Scherer, Meena Agarwal, Inoka Tennakoon, Guangnian Zhai (Dept. of Commun. Disord., 200 Health Ctr., Bowling Green State Univ., Bowling Green, OH 43403, ronalds@bgnet.bgsu.edu), Sheng Li (Xian Jiaotong Univ., Xian, P.R. China), Bogdan Kucinschi, and Kenneth De Witt (Univ. of Toledo, Toledo, OH)

This presentation summarizes results of empirical and computational studies with steady flow to indicate how normal intraglottal pressures vary as a function of (a) glottal angle, (b) radius of the glottal entrance and exit, (c) glottal obliquity, (d) inferior and superior vocal fold surface angles, and (e) false vocal fold gap. Empirically obtained intraglottal pressures differ in significant ways from earlier theoretical work and current simplified expressions. Intraglottal pressures are highly sensitive to glottal angle, entrance radius for the diverging glottis, and the exit radius for the converging glottis. Air pressures in the glottis are asymmetric when the flow is asymmetric, as they typically are in the divergent glottis. The oblique glottis produces asymmetric pressures that depend upon vocal fold angle and flow separation. Intraglottal pressures are typically lower on the side where velocities are faster. Intraglottal pressures are essentially independent of the wide range of inferior and superior vocal fold surface angles found in the human. When the ratio of the false fold gap to glottis minimal diameter is about 2, laryngeal flow resistance and intraglottal pressures reduce; flow resistance and intraglottal pressures quickly increase for gap ratios less than one. [Work supported by NIH R01DC03577.]

3aSC5. Vortex shedding and the voice source. Michael S. Howe (College of Eng., Boston Univ., 110 Cummington St., Boston, MA 02215) and Richard S. McGowan (CReSS LLC, Lexington, MA 02420)

The voice source is primarily an acoustic dipole produced by the fluctuating drag on the vocal folds [Zhao, Zhang, Frankel *et al.*, J. Acoust. Soc. Am. **112**, 2134–2146 (2002)]. In this paper the unsteady drag is determined theoretically in terms of the vorticity shed from the vocal folds. The principal source of acoustic energy is glottal-jet vorticity lying within an axial distance downstream of the glottis of less than about the glottal width. The vortex drag dipole is equivalent to the volume velocity source traditionally assumed to be located at the glottis. In addition, there exists a true, but weaker fluctuating volume source associated with volumetric changes of the vocal folds region. The relationship between the voice source dipole and sibilant fricative dipoles will be discussed. Our findings will be presented in a largely qualitative manner, so that the few equations used in the presentation can be understood physically. [Work partially supported by NIDCD-004688.]

10:00

3aSC6. Aerodynamically-based parametric description of the noise envelope in voiced fricatives. Anna Barney (ISVR, Univ. of Southampton, Southampton, SO17 1BJ, UK, ab3@soton.ac.uk) and Philip J. B. Jackson (Univ. of Surrey, Guildford, GU2 7XH, UK)

In voiced fricatives, the radiated sound is composed of a harmonic component associated with the vibrating larynx and a noise component generated at a constriction in the oral cavity. The sound from the two sources interacts in a nonlinear way to produce a noise signal with an amplitude envelope modulated at the fundamental frequency of voicing. While voiced fricatives synthesized as a linear combination of harmonic and noise components are identifiable, it is recognized that the inclusion of the modulation improves the perceived naturalness of a synthesized token. The depth of modulation of the radiated noise, for a range of aerodynamic and acoustic variables, was measured experimentally using a dynamic mechanical model of the larynx and vocal tract. Glottal excitation arose from driven shutters representing the vocal folds; frication noise was produced by an orifice plate with a sharp-edged obstacle downstream. Based on the empirical data, a parametric description was developed to predict the depth and phase of amplitude modulation of the noise from the aerodynamic and acoustic conditions.

10:20

3aSC7. Experimental investigation of the influence of a posterior commissural gap on the glottal flow resistance of a larynx model. Luc Mongeau (Dept. of Mech. Eng., McGill Univ., McDonald Eng. Bldg., 817 Sherbrooke St. West, Montreal, QC, Canada, H3A 2K6, luc.mongeau@mcgill.ca) and Jong Beom Park (Ray W. Herrick Labs., West Lafayette, IN 47907)

The influence of a posterior commissural gap (or chink) on the pulsated glottal flow was investigated using a driven physical model having a converging coronal-planar profile. Instantaneous orifice discharge coefficients of the model were obtained from the time-varying orifice area and the velocity distribution of the pulsated glottal jet measured in the axial plane using a single hot-wire probe. Sound pressures were also measured to characterize the acoustical influence of the minimum flow through the posterior chink on the radiated sound. A cyclic hysteresis of the orifice discharge coefficients was observed, which indicates a pressure head increase and a net energy transfer from the air flow onto the orifice wall. A comparison between the cases with and without the posterior chink showed that the presence of a minimal flow increased the radiated sound pressure amplitude in the frequency range in which human hearing is most sensitive.

Contributed Papers

10:40

3aSC8. A controlled-strain rheometer for the measurements of vocal fold viscoelasticity at phonatory frequencies. Roger W. Chan (Otolaryngol. Head & Neck Surg., Grad. Prog. in Biomed. Eng., Univ. of Texas Southwestern Med. Ctr., Dallas, TX 75390-9035, roger.chan@utsouthwestern.edu) and Maritza Rodriguez (Univ. of Texas Southwestern Med. Ctr., Dallas, TX 75390)

Glottal fluid-structure interactions are critically dictated by the viscoelastic properties of the vocal fold. Viscoelastic shear response of the vocal fold lamina propria has been reported previously, but with torsional rheometers data have only been obtained at low audio frequencies (up to about 80 Hz), necessitating data extrapolation to predict tissue properties in the typical phonatory frequency range. In order to characterize the viscoelastic behavior of vocal fold tissues at higher frequencies, a controlled-strain, linear simple shear rheometer was built based on the linear-motor design of the EnduraTEC ELF 3200 system. A tissue specimen was subjected to simple shear between two rigid rectangular acrylic plates, with the displacement measured by a linear variable displacement transducer (LVDT), and the shear force response detected by a piezoelectric quartz transducer. The frequency responses of the system components were assessed by accelerometers, and the complex shear modulus (G^*) of a standard ANSI 2.21 polyurethane material and that of vocal fold specimens were measured. Results showed that the rheometer can provide valid and reliable viscoelastic data of the human vocal fold cover at frequencies of up to 250 Hz. [Work supported by NIH.]

10:55

3aSC9. Multi-component synthetic model of the human larynx for investigating laryngeal fluid-structure interactions. James Drechsel, Brianne Hamilton, Adam Jepsen, Jacob Munger, Brian Pickup, Allyson Pulsipher, and Scott Thomson (Dept. of Mech. Eng., Brigham Young Univ., Provo, UT 84602)

Synthetic continuum models of the vocal folds, though only relatively recently developed, have found significant use in studying the flow-induced vibrations of the vocal folds. The advantages of these models include long lifetime and reasonable comparison with human vocal fold characteristics. However, the geometry typically employed is highly idealized, with uniformly shaped cross section and rigid mounting to rectangular plates. In this presentation, the development and characterization of a multi-component model of the human larynx is presented. The model includes the following synthetic components: multi-layer vocal folds consisting of materials with nonlinear stress-strain properties, cartilaginous and soft tissue framework, and posture control. The fabrication process, including extraction of geometric information from medical images, is summarized. Various aspects of the model are characterized to compare the model behavior with that of the human larynx. Measurements of the synthetic vocal fold material properties, including stress-strain dependence, tangent modulus, and Poissons ratio, are presented. Dependence on subglottal pressure of the model vibration frequency, flow rate, vibration amplitude, and other flow field and jet characteristics are quantified using high-speed imaging, flow visualization, and particle image velocimetry.

Session 3aSP

Signal Processing in Acoustics: Topics in Seismic Signal Processing

Max Deffenbaugh, Cochair

ExxonMobil Research and Engineering Company, 1545 Route 22, E, Annandale, NJ 08801

Alan Meyer, Cochair

665 Dunhill Dr., Danville, CA 94506

Chair's Introduction—8:00

Invited Papers

8:05

3aSP1. Observations of scattering attenuation in the Earth and implications for Q. James Rector (Dept. of Civil and Environ. Eng., UC-Berkeley, Berkeley, CA 94720)

Seismologists often attribute the attenuation of a transmitted seismic pulse to anelastic absorption or Q. Real data suggests a different dominant mechanism. The earth, particularly in sedimentary basins, is heterogeneous on all scales. This heterogeneity results in a transmitted pulse being attenuated by scattering at every boundary it encounters. Theoretical models proposed by Zien and Shapiro (1994) suggest that scattering attenuation is angle dependent, but not necessarily frequency dependent, at least for short propagation distances. In this paper we review the mechanisms for transmitted pulse attenuation both scattering- and absorption-based. We then analyze real data and compare the contributions of these mechanisms to the observed attenuation.

8:35

3aSP2. Applications of seismic interferometry to seismic field data. Roel Snieder, Kurang Mehta, Masatoshi Miyazawa, Ivan Vasconcelos (Ctr. for Wave Phenomena, Colorado School of Mines, 1500 Illinois Str., Golden, CO 80401-1887, rsnieder@mines.edu), Jon Sheiman, Rodney Calvert, Andrey Bakulin (Shell Intl. Exploration and Production, Inc., Houston, TX 77001-0481), and Anupama Venkataraman (ExxonMobil Upstream Res., Houston TX 77252-2189)

In seismic interferometry one extracts the Green's function of a system from fluctuations in the system. This makes it possible to retrieve the waves that propagate between receivers, as if one of the receivers acts as a source. Examples are presented of this principle to field data. One can show that in seismic interferometry one can retrieve the Green's function that has different boundary conditions than the physical system has. It is shown how this can be used to unravel the building response from its coupling to the subsurface, and how this principle can be exploited to suppress surface-related multiples in marine seismics. As another example, we present the propagation of P-wave and S-waves down a vertical borehole extracted from incoherent man-made noise, and how shear-wave splitting can be extracted from these measurements. Finally we show how drill-bit noise can be used for imaging when the vibrations in the drill stem have not been recorded.

Contributed Paper

9:05

3aSP3. Use of cross correlation to estimate Green's functions in the Pakistan-Himalayas. Scott D. Frank, Anna E. Foster (Dept. of Mathematics, Marist College, 3399 North Rd., Poughkeepsie, NY 12601, scott.frank@marist.edu), and Aaron N. Ferris (Weston Geophysical Corp., Lexington, MA 02420)

Recent theoretical and observational work indicates that the time derivative of the Green's function between two points can be estimated from ambient seismic noise recordings. Continuous seismograms from a temporary deployment of nine broadband seismographs in NE Pakistan (1992)

are cross correlated and averaged over 73 days to obtain Green's function estimates of the longitudinal, radial, and transverse components for all station combinations. In general, the time-averaged correlation functions show dispersion characteristics, which are analyzed using Hilbert transform filtering and wavelet transforms. Preliminary results show group velocities between 1.5 and 3.5 km/s for the period band between 2 and 20 s. In some cases, the noise correlation functions exhibit asymmetry between causal and acausal portions. This is interpreted as preferred directionality of the diffusive noise field. By comparing the peak energy at causal and acausal delay times, it appears the noise field propagates predominantly from the north-northeast during the analyzed time period.

9:20–9:35 Break

3a WED. AM

Invited Papers

9:35

3aSP4. Integrating seismic exploration methods into a geological sciences curriculum at Brigham Young Unversity. John H. McBride and R. William KeachII (Dept. of Geological Sci., Brigham Young Univ., P.O. Box 24606, Provo, UT 84602)

The continuing expansion in petroleum and other resource exploration and in geological hazard assessment for infrastructure development have combined to dramatically increase the need for training of university students in seismic imaging methods. This need is being met at Brigham Young University (BYU) by forming alliances or collaborations with private industry and government in order to provide financial support for research using seismic techniques, to obtain access to proprietary datasets, and to place students in the workplace as part of their university experience. Infrastructure support has been provided by BYU in the form of acquisition of seismic recording equipment, procuring of state-of-the-art software for data processing and geologic mapping, and building of a dedicated 3D visualization lab. This infrastructure creates an environment that mimics research and exploration programs in private industry. Seismic geophysical research foci at BYU include (1) seismic characterization of deep reservoirs for carbon sequestration, (2) 3D seismic attribute analysis for petroleum prospecting, (3) high-resolution seismic exploration applied to landslide and earthquake hazard assessments, (4) exploration of deep sedimentary basins that may be prospective for oil or gas, and (5) oilfield applications of seismic mapping in order to detect and map shallow faults that may function as leakage pathways.

10:05

3aSP5. Continuous infrasonic recording within the Utah regional seismic network. Kris Pankow, Relu Burlacu (Univ. of Utah Seismograph Stations, 135 South 1460 East, Rm. 705, Salt Lake City, UT 84112-0111, pankow@seis.utah.edu), Chris Hayward, and Brian Stump (Southern Methodist Univ., Dallas, TX 75205)

In May 2006, the University of Utah Seismograph Stations, as part of a larger consortium, installed an infrasound array east of the Salt Lake Valley (NOQ). Infrasound can be thought of as seismology in the air focusing on subaudible acoustic signals in the frequency band from 0.01 to 20 Hz. Advantages of installing an infrasound array in northern Utah include a proximity to solid earth sources (mine blasts and earthquakes) and the possibility of incorporating the data into a regional seismic network. The primary application of infrasound is in identification of explosions detonated at or near the solid earth-atmosphere boundary. However, to fully utilize the infrasonic data, better understanding of infrasonic source characteristics and propagation is needed. Goals of the infrasound installation at NOQ include (1) use of local ground truth information to assess the role of infrasound in source identification; (2) classifying properties of the infrasound signal by source type; (3) assessing the time varying characteristics of infrasound propagation; and (4) incorporating the infrasound data into seismic source location. In this paper, we will show examples of both seismic and infrasound signals from ground truth events and we will discuss aspects of data processing.

Contributed Papers

10:35

3aSP6. Modeling the spatial coherence of seismic surface waves at short propagation distances. D. Keith Wilson (U.S. Army Engineer Res. Dev. Ctr., 72 Lyme Rd., Hanover, NH 03755-1290, D.Keith.Wilson@erdc.usace.army.mil) and Vladimir E. Ostashev (NOAA Earth System Boulder, CO 80305, New Mexico State Univ., Las Cruces, NM 88003)

Scattering from small, random inhomogeneities in the ground (such as rocks and voids) diminishes the spatial coherence of seismic waves, much as turbulence in the atmosphere diminishes spatial coherence of propagating sound waves. In this paper, modeling of seismic spatial coherence for surface waves propagating over short distances (roughly 50 m to 1 km) and at relatively high frequencies (greater than approximately 1 Hz) is considered. The random inhomogeneities are distributed broadly in size and are described with a quasi-wavelet (QW) model. For simplicity, we consider propagation of a 2-D scalar wavefield. Scattering from an individual QW is first calculated using a single-scattering approximation. Then, the contributions from all randomly placed QWs, including all size classes, are cross-correlated to find the spatial coherence of the total wavefield. As an example, we consider modeling of seismic coherence at a site in the Mojave Desert with random basaltic inhomogeneities surrounded by sand. The impact of coherence loss on direction-finding with geophone arrays is discussed.

10:50

3aSP7. Numerical simulation of rock acoustics with a multi-phase model. Limin Song (ExxonMobil Res. and Eng. Co., 1545 Rte. 22 E., Annandale, NJ 08801), Yang Zhang (MIT, Cambridge, MA 02139), and Max Deffenbaugh (ExxonMobil Res. and Eng. Co., Annandale, NJ 08801)

Understanding how the effective acoustic properties of porous rock are related to its physical properties is very important in exploration and production of oil and gas, and remains an active research topic. We present a computational method to study the acoustics of fluid-filled porous rock over a range of frequencies. The method solves the elastic wave equation in the solid rock matrix and the viscous wave equation in the pore fluid. The coupled multi-phase wave equations are solved numerically over a digital representation of the rock sample where the material phase and properties at each volume cell are identified. The significance of the new method is that the viscous effects of pore fluid on the bulk acoustic response are modeled accurately with the linearized Navier-Stokes equation instead of with the generalized Maxwell body low frequency approximation that is commonly used to overcome the difficulty of modeling frequency-dependent fluid shear modulus in time domain. The rotated staggered grid finite difference scheme is used to solve the coupled multi-phase wave equations. We simulate P-wave propagation through an idealized rock model of alternating solid and fluid layers where an analytical solution is available and obtain excellent agreement between numerical and analytical results.

11:05

3aSP8. Mapping the middle Paleozoic through three-dimensional seismic attributes in the central Arkoma Basin, Oklahoma. Alonzo R. Brinkerhoff, Bill Keach, John McBride, Scott Ritter (Brigham Young Univ., ESC, Provo, UT 84602), and Vince Felt (BP America Production Co., Houston, TX 77253)

Seismic attribute analysis and mapping, reveals that Hunton Group strata do occur in the central and southern Arkoma Basin of eastern Oklahoma and that they were not completely removed by pre-Woodford erosion as previously believed. Well and isochron data through the Viola-Woodford interval (Ordovician-Late Devonian) show isolated ~40 m thick lenses of Hunton rocks, on average measuring 3 km in diameter. This distribution could be explained in two ways; 1) Hunton occurrences could represent isolated erosional remnants reflecting incomplete removal of the Hunton Group during Middle Devonian time (pre-Woodford unconformity) or 2) due to karsting and collapse of stratigraphically lower units (Arbuckle Group?), lenses of Hunton rocks could have sagged into sinkholes where they were preserved below regional base level. A combination of these models may be necessary to account for areas where we see thinning both in the Woodford and Viola, suggesting that a Hunton lens is structurally lowered due to karsting, but due to its erosionally resistive nature, the lens forms a depositional high, causing the Woodford to thin over it.

WEDNESDAY MORNING, 6 JUNE 2007　　　　　　　　　　　　　　CANYON ROOM B, 8:00 TO 11:35 A.M.

Session 3aUW

Underwater Acoustics: Environmental Uncertainty and Characterization

Marcia J. Isakson, Chair
Univ. of Texas at Austin, Applied Research Lab., 10000 Burnet Rd., Austin, TX 78758

Chair's Introduction—8:00

Contributed Papers

8:05

3aUW1. Effects of model discretization on the statistics of model parameters in nonlinear inverse problems. Andrew A. Ganse, Robert I. Odom (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Andrew A. Ganse, and Robert I. Odom (Univ. of Washington, Seattle, WA 98195)

An inverse problem is the estimation of a continuous model function given a finite set of data points. In most geophysical problems, due to data noise and lack of complete geometric coverage, this problem is intractable analytically so we parameterize the continuous model and solve as a regularized parameter estimation problem. Coarsening the discretization of an ocean bottom model is one way to stabilize the otherwise ill-conditioned problem; this is sometimes called regularization by discretization. But by setting the resolution of the model estimate prior to the inversion in this way, one also affects the model estimate's statistics, due to the inherent tradeoff between model resolution and uncertainty. This relationship is well understood for linear problems, but resolution and uncertainty in nonlinear inverse problems remains an active area of research. In nonlinear problems the probability distributions of the model parameters are not Gaussian, and so with Monte Carlo methods we explore the relationship between the magnitude of non-Gaussian features in the model probability, the discretization of the model, and the aliasing of its parameters. [Work partially supported by ONR.]

8:20

3aUW2. Experimental design for acoustic characterization of a seismic exploration 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 (Univ. of Louisiana, Lafayette, LA 70504), Joal J. Newcomb (Naval Res. Lab., Stennis Space Ctr., MS 39529), James M. Stephens, and Grayson H. Rayborn (Univ. of Southern Mississippi, Hattiesburg, MS 39406)

The Littoral Acoustic Demonstration Center (LADC) (participating institutions given in author list) is scheduled to perform an experiment in the Gulf of Mexico to characterize the underwater acoustic field of a seismic exploration airgun array. The design is constrained by the array lengths and hydrophone spacings of the ten four-channel LADC calibrated environmental acoustic recording system (EARS) buoys available for the measurements. Placement of the EARS arrays is determined to achieve best the design goals, given the constraints. The fixed mooring EARS hydrophone depths will vary from 100 to 1171 m, and near-surface ship-deployed phones will be used for shallower depths. Straight-line source boat paths of varying length and spacing have been calculated that redundantly fill all solid angle bins whose size is determined by small uncertainties in the array location and orientation [Robert Laws (2006), private communication]. Shot spacing is also a factor. Care is taken to have nearby shots to the EARS arrays for all bins. All direct paths are calculated including refraction. Data are presented in spherical and planar images to show angular and bin coverages and shot ranges. [Work supported in part by the Joint Industry Project through the International Association of Oil and Gas Producers.]

8:35

3aUW3. Benchmarking echoes from an object in an ocean waveguide with bottom roughness. Steven A. Stotts (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, stotts@arlut.utexas.edu)

A complex k-plane two-way coupled mode model including both trapped and continuum effects can be used to benchmark acoustic reverberation in an ocean waveguide. In addition, scattering from both bottom roughness and an object in the waveguide can be evaluated simultaneously in a consistent approach. To derive the benchmark, a two-dimensional problem for the recent Reverberation Workshop, held in Austin, TX in November 2006, was selected. The problem has a constant sound speed Pekeris waveguide and is used as the waveguide model for developing the current benchmark. The two-dimensional problem is derived from a vertical slice through a three-dimensional environment and only considers scattering in the vertical plane. The source is a Gaussian beam with a center frequency of 250 Hz. The source and receiver are mono-static in range and the waveguide has a mean depth of 50 m and includes bottom roughness obtained from a single realization of a defined roughness spectrum. For the benchmark calculation, an object (in addition to the bottom roughness) is introduced into the water column. Propagation of the signal to and echo from the object in the water column also includes bottom scattering effects.

3aUW4. Suspended-sediment measurements in a fluvial environment using acoustic backscatter. Wayne O. Carpenter, Jr., James P. Chambers, Christopher K. Smith (Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, wocarpen@olemiss.edu), Daniel G. Wren, and Roger A. Kuhnle (USDA-ARS-NSL, Oxford, MS 38655)

The continued development of acoustic backscatter technology for suspended-sediment measurement is an important goal that, when attained, will result in sediment transport data of unprecedented quality and quantity. Towards this end, an instrumented raft has been deployed in an ephemeral stream in the North Mississippi area. The raft is used as a floating instrument platform that allows acoustic transducers to stay in contact with the water surface during stage changes caused by runoff into the stream channel. The single-frequency acoustic data is calibrated for both particle size and concentration using concurrent pumped samples of the water/sediment mixture collected from various levels above the stream bed. Data from several storm events has been acquired and will be presented, including time series of converted suspended-sediment concentration data and estimates of errors introduced during the conversion process.

3aUW5. Seabed characterization and transmission loss estimation from geoacoustic inversion using a towed array. Yong Han Goh, Chen-Fen Huang, Peter Gerstoft, and William S. Hodgkiss (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0238, yhgoh@ucsd.edu)

The goal of geoacoustic inversion is to estimate environmental parameters from measured acoustic fields from, e.g., a towed array. Though the inversion results have some uncertainty, inversion is an efficient technique to estimate environment parameters. Based on the *a posteriori* probability density of environment parameters obtained from inversion, we perform statistical estimation of transmission loss (TL) and generate a 90% credibility level envelope or uncertainty band for the TL. This uncertainty band accounts for the inherent variability of the environment not usually contained in sonar performance prediction model inputs, and presents a useful probabilistic description of the environments variability. Towed arrays are advantageous to fixed systems as they are easy to deploy and the moving ship enables estimation of spatially variable seabed properties. The approach is demonstrated with data obtained from the MAPEX2000 experiment conducted by NATO Undersea Research Centre (NURC) in the Mediterranean Sea in November 2000.

3aUW6. Sediment attenuation effects in sound transmission results from the New Jersey Shelf. Simona M. Dediu, William L. Siegmann (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY, 12180), William M. Carey, and Allan D. Pierce (Boston Univ., Boston, MA 02215)

Experiments were conducted near the site of AMCOR Borehole 6010 on the New Jersey Shelf to characterize propagation predictability. The importance of a non-linear power-law frequency dependence of the sediment volume attenuation in the uppermost sediment layer is demonstrated. One metric of transmission loss variation with range is an effective attenuation coefficient that can be extracted from measurements and calculated with the parabolic equation. Previously it was found [W. M. Carey and R. Evans, J. Oceanic Eng., **23**] that a power exponent of 1.5 modeled the measurements. The present approach uses 1 kHz as an attenuation reference frequency and employs different parameter ranges and optimization criteria. For 400–1000 Hz, this procedure leads to a power exponent in the range 1.7–2.0, which is consistent with other sand-silt regions. The estimates are robust with respect to variations in the water and sediment sound-speed profiles and the sediment layer thickness. The influence of measured range dependence in the sound speed and bathymetry is examined. Estimates of signal time spread are calculated, which can also be obtained for recent experiments in the same ocean region. [Work partially supported by ONR]

3aUW7. Investigation of three-dimensional propagation effects at the New Jersey shelf break front. Kristy A. Moore, James H. Miller, Gopu R. Potty (Dept. of Ocean Eng., Univ. of Rhode Island, Sheets Bldg., Narragansett Bay Campus, Narragansett, RI 02882), James F. Lynch, and Arthur Newhall (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Signals recorded on the WHOI VLA/HLA from a ship-towed J15 source during the Shallow Water Experiment (SW06) in the New Jersey Shelf were analyzed for three-dimensional (3D) propagation effects. The signal was a simple CW tone at 93 Hz and the source depth was approximately 50 meters. This 3D propagation effect is a horizontal analogue to the classical Lloyds mirror effect, where fronts can totally internally reflect sound incident upon them at low grazing angles. The direct and reflected modal rays can constructively interfere, having the potential to increase the intensity level by 6dB [Lynch *et al.*, IEEE J. Ocean. Eng., **31** 33–48 (2006)]. Trapping of sound between the shelf break front and an internal wave packet may reduce transmission loss by 10 to 20 dB. Bathymetric refraction is also a possible contributor to the reduced transmission loss. Modeling of the Lloyds mirror effect was carried out using the Kraken normal mode model with the environmental parameters of the New Jersey Shelf and J15 source. Analysis of the acoustic recordings of the J15 source and ship noise from the R/V Knorr is consistent with the modeling results. [Work supported by Office of Naval Research.]

3aUW8. Determination of frequency dependence of sediment attenuation in shallow water using the combustive sound source. David P. Knobles, Preston S. Wilson, and Steve E. Cho (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Acoustic signals generated by the combustive sound source (CSS), an impulsive source based on hydrogen-oxygen combustion, were recorded at various source-array ranges on two L-arrays during the Shallow Water'06 experiment off the coast of New Jersey, in about 70 m of water. The bandwidth of the CSS is about 3 kHz and thus offers an opportunity to determine those geo-acoustic parameters of the seabed that are frequency dependent. One parameter that has received significant attention is the sediment attenuation due to its importance in correctly predicting transmission loss in shallow water. In principle, if one can determine the real part of the sediment sound speed and layering structure, the received acoustic data over a 3 kHz band could be used in combination with an advanced propagation model to determine the frequency dependence of the sediment attenuation. The central hypothesis of the study is that while such an inverse approach is subject to non-uniqueness, a significant number of physical measurements such as the sound speed profile in the water column, bathymetry, and sub-bottom layering that were made along the propagation paths may provide adequate constraints that allow for a unique determination of the sediment attenuation. [Work supported by ONR.]

3aUW9. Energy and vorticity analysis (EVA) experiment 2006. Deep Ghosh and Henrik Schmidt (MIT, 77 Massachusetts Ave., #5-204, Cambridge, MA 02139)

Low frequency (2–15 kHz) monostatic and multistatic target scattering data was acquired at subcritical and supercritical angles from three targets: a composite cylinder, a composite sphere and an air filled GOATS sphere. A TOPAS parametric source with 40 kHz primary and 2–15 kHz secondary frequency was used to insonify the targets for the monostatic and multistatic near field measurements. The monostatic experiment was carried out using a 16 element Vertical Line Array (8 kHz). The multistatic experiment was carried out using a CW Doppler source at 17 kHz with a 4 element receiver array deployed from the ships well. Differential GPS data was recorded for accurate track recording. The ship roll, pitch and heading were also recorded. The multistatic near field measurements were

carried out using a light PVC dome of height 1.7 m and diameter 3.7 m. Nine broadband hydrophones served as the receiver array which could be rotated by an electric motor. Aspect of -145 degrees to $+145$ degrees was covered in each case. The data will be processed to study the near field scattering from targets which would help in better target classification and detection in near field region. [Work supported by NURC, NRL and Leonardo.]

10:35

3aUW10. *In situ* reflection coefficient measurements from 5–50 kHz from the experiments for validation of acoustic modeling techniques (EVA) sea test. Marcia Isakson, Nicholas Chotiros, James Piper (App. Res. Labs., The Univ. of Texas, P.O. Box 8029, Austin, TX 78713), and Mario Zampolli (NATO Undersea Res. Ctr., La Spezia, Italy)

An accurate model of high-frequency reflection loss at the ocean bottom is critical to understanding shallow water sound propagation and the detection and classification of buried objects such as mines. The value of the observed reflection coefficient can be perturbed by such effects as scattering from a rough interface and dispersion in the ocean sediment. To obtain ground truth measurements for model development and verification, reflection coefficients from a rough water/sediment interface were measured as part of the experiments for validation of acoustic modeling techniques (EVA) sea test on the north shore of Isola dElba, Italy. Data at grazing angles from 10–80 deg and frequencies from 5–50 kHz were obtained. Experimental effects such as transducer response and spherical wave effects were quantified. The data will be analyzed to determine the influence of dispersion and scattering effects. [Work supported by ONR, Ocean Acoustics.]

10:50

3aUW11. Bayesian focalization and tracking. Stan E. Dosso and Michael J. Wilmut (School of Earth and Ocean Sci., Univ. of Victoria, Victoria BC, Canada)

This paper describes a Bayesian approach to source localization and tracking with uncertain ocean environmental parameters (water column and seabed). The inversion is formulated for both source location and environmental parameters, and solved using Gibbs sampling methods which sample directly from the posterior probability density. The information content for source localization is quantified in terms of probability ambiguity surfaces, consisting of joint marginal probability distributions for source range and depth integrated over unknown environmental parameters, similar to the optimum uncertain field processor (OUFP). The posterior environmental information content, including both prior information and information provided by the acoustic data, is also quantified in terms of marginal distributions. In source tracking, knowledge of the environmental parameters can be augmented as the tracking progresses, by explicitly passing posterior environmental information from one localiza-

tion on, as improved prior information in the subsequent localization. Information on interparameter correlations is included by formulating marginal distributions in terms of principle components (empirical orthogonal functions) of the environmental parameters.

11:05

3aUW12. Temporal variability of narrow-band tones in very shallow coastal waveguides. Stephen D. Lynch and Gerald L. D'Spain (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California—San Diego, La Jolla, CA 92093-0701)

During an experiment near San Clemente, CA, a source moored on the bottom just outside the surf zone (500 m offshore, approximately 9-m depth) transmitted eight tones between 70 and 700 Hz to two nearly perpendicular, 64-element bottom-lying hydrophone arrays located approximately 1.2 km downcoast and 1 km further offshore (approximately 12-m depth). The acoustic tones were transmitted for 5 min every half hour over a 22-h period. Surface wave data obtained from a depth sensor array located approximately 40 km upcoast, as well as data from a PUV sensor data deployed during the experiment, are used to compute concurrent directional surface wave spectra. Long fast Fourier transforms of the acoustic recordings reveal temporal and spatial variability of the width and shape of the received tones in temporal frequency space. Surface-wave-related mechanisms of acoustic frequency distortion are proposed and their relative significance is investigated through the comparison of the data with results from physics-based models that use the measured surface wave properties as input parameters.

11:20

3aUW13. Trends over minutes to decades in oceanic ambient sound measured off the southern Californian coast. Rex K. Andrew, Charlotte V. Leigh, Bruce M. Howe, and James A. Mercer (Appl. Phys. Lab., 1013 NE 40th St., Seattle WA 98105)

Ambient sound spectra have been collected since 1994 from a decommissioned deep water military acoustic receiver located southwest of San Nicolas Island (southern California). A new spectrum is collected every 6 minutes. The dataset can be used to study level variability on scales from minutes to years. The calibration curves for this system, however, were considered suspect. Recently, a calibrated acoustic recorder was deployed near the receiver (McDonald *et al.*, JASA, **120**(2), August 2006). Comparison of spectra for December 2003 provides a long-sought-after modern absolute calibration. With this new correction, the dataset can additionally be compared to levels measured in the 1960's by the same receiver. Although the corrected measurements corroborate (albeit by construction) the 1960–2000 ambient noise increase reported by McDonald *et al.*, there is scant evidence of an increase over the last decade. This study again highlights the need for on-going long-term well-calibrated observation programs. [Work sponsored by ONR.]

Meeting of Accredited Standards Committee (ASC) S1 Acoustics

J. P. Seiler, Chair S1

U.S. Department of Labor, Mine Safety and Health Admin., P.O. Box 18233, Bldg. 38, Cochrans Mill Road, Pittsburgh PA 15236

G. S. K. Wong, Vice Chair S1

Institute for National Measurement Standards, National Research Council, Montreal Road, Bldg. M36, Ottawa, Ontario K1A OR6, Canada

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 10:30 a.m. on Tuesday, 5 June 2007.

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 S3

University of Texas, Department of Communication Sciences & Disorders, CMA 2-200, Austin, TX 78712

R. F. Burkard, Vice Chair S3

Hearing Research Laboratory, State University of New York at Buffalo, 215 Parker Hall, Buffalo, New York 14214

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 10:30 a.m. on Tuesday, 5 June 2007.

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.

Session 3pAA

Architectural Acoustics: Acoustics in Rooms, Ducts, and Forensics

Alexander U. Case, Chair

Fermata Audio and Acoustics, P. O. Box 1161, Portsmouth, NH 03802-1161

Chair's Introduction—1:00

Contributed Papers

1:05

3pAA1. An evaluation of acoustical characteristics of classrooms in tropical environments. Ramachandraiah Alur and Nithya Subramaniam (Dept. of Civil Eng., IIT-Madras, Chennai-36, India, ram_alur@yahoo.com)

Classrooms in India, especially those located in warm humid areas, are generally equipped with mechanical ventilation such as fans or air-conditioning systems which influence the background noise of these classrooms. These systems obviously lead to a high level of background noise even during unoccupied conditions. Ventilation requirements become acute in summer and rooms are provided with a lot of open windows. This will be a problem if there are high levels of external noise. An acoustical analysis has been carried out to assess the acoustical quality and conditions for speech communication. Speech intelligibility tests along with physical and acoustical measurements were made in various classrooms (IIT Madras campus). Octave band measurements of background noise levels and reverberation times were made. The measured values have been compared with a few existing standards such as ANSI. Preliminary surveys have been administered to students and instructors to get a subjective opinion of their experiences in these environments. The classrooms consisted of untreated, partially treated, and fully treated rooms. The inferences from this study have been drawn to suggest better design procedures for such classroom environments.

1:20

3pAA2. Utilizing variable acoustics at Snow College's Jorgensen Concert Hall—A user's perspective along with acoustical parameters measured in the hall. Vance E. Larsen (Horne School of Music, Snow College, 150 E. College Ave., Ephraim, UT 84627, Vance.Larsen@snow.edu) and Bill Dohn (Dohn and Assoc., Inc., Morro Bay, CA 93442)

The Jorgensen Concert Hall is the centerpiece of the Horne School of Music at Snow College in Ephraim, Utah. It seats about 720 people and was completed in 2004. By adjusting reverberation and sound reflections on the stage and in the house, the Jorgensen accommodates a wide variety of college concerts, including jazz big band, chamber ensembles, choral groups, concert band, and symphony orchestra. The Jorgensen has also become popular with visiting groups such as the Utah Symphony, the Orchestra at Temple Square, and the Julliard Jazz Trio. Successes and challenges in using the Jorgensen will be discussed, along with a presentation of acoustical parameters measured in the hall.

1:35

3pAA3. On the acoustics of Hellenistic amphitheaters. Nico F. Declercq and Cindy S. A. Dekeyser (Woodruff School of Mech. Eng., GATech, 801 Ferst Dr., Atlanta, GA 30332-0405, and GT-Lorraine, 2 r. Marconi, 57070 Metz, France, nico.declercq@me.gatech.edu)

Hellenistic theaters are well known for their marvelous acoustics. A numerical study reveals that the seat rows of the theater unexpectedly play an essential role in the acoustics, at least when the theater is not fully filled

with spectators. The seats, which constitute a corrugated surface, serve as an acoustic filter that passes sound coming from the stage at the expense of surrounding acoustic noise. Whether a coincidence or not, the famous theater of Epidaurus was built with optimized shape and dimensions.

1:50

3pAA4. Archaeoacoustics re-examined. Matthew C. M. Wright (ISVR, Univ. of Southampton, Southampton, SO17 1BJ, UK, mcmw@soton.ac.uk)

Archaeoacoustics examines the acoustic properties of ancient artifacts, buildings and natural spaces in order to shed light on the past. In one such study by Jahn, Devereux, and Ibison [J. Acoust. Soc. Am. **99**, 649–658 (1996)] it was found that a number of ancient enclosures (principally neolithic burial chambers) exhibited similar resonance frequencies. It has subsequently been suggested that their results imply that these chambers were used for chanting and were intentionally built to enhance this. A re-analysis of their data will be presented which indicates that this is unlikely to be a valid conclusion. [This research was supported by an EPSRC Advanced Research Fellowship.]

2:05

3pAA5. An acoustic raindrop effect at Chichen Itza in Mexico. Jorge Antonio Cruz Calleja (Esc. Sup. de Ing. Mec. y Electr. UC, Dept. of Acoust., Av. Sta Ana 1000 Mexico D.F. Del. Coyoacan, C.P. 04430, San Francisco Culhuacan, Mexico) and Nico F. Declercq (GATech, Atlanta, GA 30332-0405)

One of the acoustic effects caused by the staircase of the El Castillo pyramid at Chichen Itza is the so-called raindrop effect [Declercq *et al.*, J. Acoust. Soc. Am. **116**(6), 3328–3335 (2004)]. When visitors climb the colossal staircase of Maya pyramids, their footsteps are transformed into sound having distinct frequencies similar to raindrops falling in a bucket filled with water. *In situ* experiments are reported followed by numerical simulations of the raindrop effect together with a physical explanation. In addition to numerical simulations, a rule of thumb formula is extracted from the calculations that enable the prediction of the acoustic raindrop frequency at any other pyramid in Mexico. If the raindrop effect is a phenomenon that was intentionally incorporated in the construction of the Maya pyramids, such as the pyramid in Chichen Itza, then it was most probably related to the rain god Chac for which there is ubiquitous archaeological evidence decorated on the pyramid itself.

2:20

3pAA6. Sound attenuation in ducts utilizing spatially periodic area changes with absorbing material. Lisa Burton, Donald Bliss, and Linda Franzoni (Dept. of Mech. Eng., Duke Univ., Durham, NC 27708)

The propagation of quasi-one-dimensional acoustic waves through a duct with spatially periodic area changes has been studied theoretically and experimentally in previous research. This configuration exhibits stop-band and passband behavior, with substantially reduced sound transmission in the stop bands, but little effect in passbands. In the present study

3129 J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America **3129**

3p WED. PM

the periodically spaced regions of the duct with larger cross-sectional area are partially filled with porous material, leaving a constant area passage for airflow. The effect of the porous material is to make the propagation wave number in the larger cross-sectional areas complex, thereby providing significant attenuation even in passbands. To determine this wave number, a fairly simple theory based on cross-sectional averaging and a higher order correction dependent on geometry is developed to achieve excellent accuracy. The dispersion relation for the duct and the transfer function for the overall system are analyzed using a transfer matrix approach. The relationship between area change parameters and the location and width of stop bands is discussed. Experimental measurements for ducts with periodic area changes having open cell foam as the porous material are compared with the theory. Application to a muffler designed to place target engine harmonics in stop bands is shown.

2:35

3pAA7. Prediction of porous media properties using a cross-sectional averaging theory for one-dimensional sound propagation in ducts having two propagation media. Lisa Burton, Donald Bliss, and Linda Franzoni (Mech. Eng., Duke Univ., Durham, NC 27708)

An alternative method to measure the complex effective wave speed and density of bulk reacting porous materials is described. A hybrid duct partially filled with porous material, having constant cross-sectional properties, is studied theoretically and experimentally. A fairly simple quasi one-dimensional theory based on cross-sectional averaging is derived, compared to exact solutions, and found to work extremely well below the cut-on frequency for higher modes. The basic theory depends only on the ratio of cross-sectional areas and the properties of the individual propagation media, but not on the configuration of material in a cross-section. Accuracy can be improved at higher frequencies through a higher order correction that includes a shape coefficient that depends on the specific cross-sectional configuration. Once the propagation wavenumbers are known for two or more cross-sectional configurations having the same two

media, e.g. air and porous material, it is shown that the porous material properties can be extracted relatively easily. Experimentally, porous material samples line the sides of a tube closed at one end, and the complex wavenumber is determined from standing wave measurements. Results are compared for various lining configurations to assess the accuracy of the method and the sensitivity to cross-sectional configuration in practice.

2:50

3pAA8. Methodology for reconstruction of vehicle accident acoustics for use in forensic visualization. William Neale and Toby Terpstra (Kineticorp LLC, 44 Cook St., Ste. 510, Denver, CO 80206, wneale@kineticorp.com)

This paper describes a methodology for incorporating physically accurate sound composition in forensic visualizations. The use of sound in forensic visualization provides the viewer a more realistic and comprehensive understanding of actual accident events. Forensic visualization can represent complex events such as car crashes through animation, making it easier to understand the accident. Without sound, however, a visual representation of an accident will lack important information. Sound adds a spatial dimension for the viewer since sound defines space through such properties as reflection and reverberation. In addition, sound provides an understanding of important details of the accident such as the duration and severity of an accident, and potential audible sounds heard by a witness. Sound also allows the viewer to experience events in the accident that are outside the viewer area or occluded from view. Currently, there is no methodology for compositing sound in an animation to follow the principles of sound and reflect the specifics of an accident. Acoustical principles define how sound attenuates, reflects, dampens, blends, and changes in pitch and sound level. The unique circumstances of an accident define what sounds are present and the timing and sequencing of these sounds. This paper provides a methodology for creating sound that both follows acoustical principles and reflects the unique circumstances of an accident.

WEDNESDAY AFTERNOON, 6 JUNE 2007

ALPINE EAST, 1:00 TO 2:00 P.M.

Session 3pAO

Acoustical Oceanography: Acoustical Oceanography Prize Lecture

Bruce M. Howe, Chair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

Chair's Introduction—1:00

Invited Paper

1:05

3pAO1. The recent history of our understanding of low-mode internal tides in the ocean. Brian D. Dushaw (Appl. Phys. Lab, Univ. of Washington, 1013 N.E. 40th St., Seattle, WA 98105, dushaw@apl.washington.edu)

The past decade has seen a renaissance in the interest in oceanic internal tides and their role in mixing the abyssal ocean. One of the sparks of this renaissance was the discovery of the remarkable coherence of low-mode internal tides using an ocean acoustic tomography array deployed in the central North Pacific Ocean in 1987. The result was subsequently confirmed by satellite altimetry. This talk will review these results and related results from the 1991 Acoustic Mid-Ocean Dynamics Experiment (AMODE) in the North Atlantic and the 2001 Hawaiian Ocean Mixing Experiment (HOME) around the Hawaiian Ridge. Internal tides are generated by tidal forces hence they are a means by which the tides dissipate energy. With wavelengths of 150-km, low-mode semidiurnal internal tides appear to propagate 1000's of kilometers across ocean basins while retaining considerable coherence.

Session 3pED

Education in Acoustics: Acoustics Education Prize Lecture

Uwe J. Hansen, Chair
64 Heritage Dr., Terre Haute, IN 47803

Chair's Introduction—2:15

2:20

3pED1. Descriptive acoustics of music and speech. William J. Strong (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602, william_strong@byu.edu)

A brief background will be given of "Descriptive Acoustics of Music and Speech," an introductory course taught at Brigham Young University for the past 40 years. Several conceptual and physical models used in the course will be described. Some phenomena observed in music and speech will be explained in terms of the models. Diagrams, sound samples, and video clips will be used to illustrate the phenomena and their explanation.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Barbara G. Shinn-Cunningham, Chair
Boston Univ., 677 Beacon St., Boston, MA 02215

Chair's Introduction—1:05

1:10

3pID1. Hot topics in animal bioacoustics. David Mann (Univ. of South Florida, College of Marine Sci., 140 7th Ave. S., St. Petersburg, FL 33701, dmann@marine.usf.edu)

Discoveries in animal bioacoustics continue to take advantage of new technology and the integration of field and laboratory studies. Zuk and colleagues described the evolution of a silent cricket morph in response to selection from an acoustically orienting parasitoid fly on the Hawaiian Island of Kauai. Feng, Narins, and colleagues have demonstrated ultrasonic communication in the concave-eared torrent frog using field playback experiments and laboratory neurophysiological studies of the auditory system. Field research using sound traps has suggested that some pelagic larval fishes use sound to locate coral reefs, yet laboratory studies suggest that larval fish hearing may not be sensitive enough to detect reefs from large distances. Several groups have used evoked potentials to study hearing in diverse species of odontocetes. The grand challenge for the future is hearing measurements from mysticetes. Tyack, Madsen, Johnson, and colleagues have used animal-borne acoustic recording tags to study the feeding behavior of deep-diving beaked whales by analyzing outgoing echolocation pulses and their incoming returns from prey targets. The McCauley Library at Cornell University has developed a powerful web-based interface to their extensive database of animal sounds (www.animalbehaviorarchive.org). This system provides a simple, powerful tool for future comparative animal bioacoustics studies.

1:30

3pID2. Rehabilitating the world's worst sounds: Time-frequency reassignment in sound modeling and morphing. Kelly Fitz (Starkey Hearing Res. Ctr., 2150 Shattuck Ave., Ste. 408, Berkeley, CA 94704, kelly_fitz@starkey.com)

Sound morphing implies a gradual transformation from one sound to another, or the transfer of features of one sound to another. Morphing has been used to synthesize the voices of aliens and animals and a wide variety of sound effects in films and popular music. Additive sound models provide a flexible medium for performing sound morphing and manipulation. The application of time-frequency reassignment—the computation of precise, signal-dependent loci for time-frequency analysis data from first-order partial derivatives of short-time spectral phase—to the construction of robust, manipulable additive sound models will be discussed, and the advantages afforded by the increased precision of reassigned sound models will be demonstrated. Machines, mammals, and musical instruments will be morphed and manipulated, and an attempt will be made to rehabilitate the very worst sounds in the world.

3pID3. Noise and its hot topics. Brigitte Schulte-Fortkamp (TU-Berlin, Einsteinufer 25, 10587 Berlin, Germany)

Noise is an increasing challenge; the pollution of the environment comprises a multitude of risks. Fortunately, there is also great effort for assessment and evaluation that touches and also includes socio-cultural, psychological, architectural, ecological, and health fields. To sufficiently consider local and regional conditions there is an ongoing process within TC Noise to understand soundscapes as a new tool for the measurement, analysis, and design of environmental and community noise that applies the knowledge of both science and community experts. The process for standardization is under initiative. Classroom acoustics, which achieved great success, is now on the way to enforce its widespread acceptance supported by members from TC Noise, Architectural Acoustics, Speech Communication, ASA-Committee on Standards, and others. Work on improved noise control continues and places emphasis on noise in hospitals. The importance of noise in health care suggests that the problem of hospital noise is universal, and that noise control techniques might also be expected to be applicable broadly. Taking into account the fact that the life quality of millions of people is severely impaired by noise from different sources, we will keep moving.

WEDNESDAY AFTERNOON, 6 JUNE 2007　　　　　　　　　　　　　　TOPAZ ROOM, 1:00 TO 3:05 P.M.

Session 3pPP

Psychological and Physiological Acoustics: Masking and Temporal Processing

Jungmee Lee, Chair

Univ. of Arizona, Speech, Language and Hearing Science, P. O. Box 210071, Tucson, AZ 85721-0071

Chair's Introduction—1:00

Contributed Papers

1:05

3pPP1. Evidence that stimulus bandwidth affects temporal resolution and temporal integration in naive listeners. Julia A. Mossbridge and Beverly A. Wright (Dept. of Comm. Sci. and Disord., 2240 N. Campus Dr., Evanston, IL 60208, j-mossbridge@northwestern.edu)

The separation of auditory events in time (temporal resolution) and the summation of these events over time (temporal integration) are both essential to normal hearing. Here, it is reported that both of these abilities are affected by stimulus bandwidth in naive listeners. In a single group of 30 naive listeners, amplitude-modulation detection thresholds were determined over a range of modulation rates, and signal-detection thresholds over a range of signal durations, using noises of two different bandwidths (0.25 kHz and 3.80 kHz), both with a 4.40-kHz upper cutoff frequency. Amplitude-modulation thresholds were adjusted mathematically to compensate for reductions in modulation depth introduced by filtering after modulation. The average time constants estimated from these data were significantly larger for the narrowband than the broadband noise for both temporal resolution (5.1 versus 1.3 ms; p<0.0001) and temporal integration (251 versus 13 ms; p<0.0001). Thus, naive listeners showed shorter temporal resolution and integration windows with a wider stimulus bandwidth, suggesting that, at least for this population, both of these temporal-processing abilities are influenced by the spectral parameters of the input. [Work supported by a grant from NIH/NIDCD.]

1:20

3pPP2. Using method of adjustment to measure informational masking using multitone maskers. William Yost, Kristine Cabael, and Stanley Sheft (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sherdian Rd., Chicago, IL 60626, wyost@luc.edu)

Previous research has shown that the method of adjustment (MoA) procedure can produce masking results that differ from those obtained with force-choice (FC) procedures in that the MoA procedure produces less informational masking in some contexts and less backward interference masking than FC procedures. A comparison between MoA and FC procedures has not been tested using a random, multimasker condition often used to study informational masking. The present study compared a MoA and a same–different FC procedure using a random 1, 10, or 100 sinusoidal-component masker or a noise masker for a 2000-Hz signal (nominal masker bandwidth was 800 to 6000 Hz). Two groups of 20 naive subjects each were tested. In general, the results indicate that the mean data for the MoA and FC procedures did not differ, but the intersubject variability did, with the MoA procedure yielding smaller variability. The results will be discussed in terms of attention and its relationship to informational masking. [Work supported by grants from NIDCD.]

1:35

3pPP3. Spatial release from speech-on-speech masking in adult hearing aid users. Nicole L. Marrone, Christine R. Mason, and Gerald Kidd, Jr. (Dept. of Speech, Lang., and Hearing Sci., Hearing Res. Ctr., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215)

Although it is well documented that spatially separating speech maskers from a target speech source normally results in a release from masking, little is known about this effect in hearing aid users. The current study addressed the interaction between spatial release from masking, reverberation, age, and amplification in a multitalker soundfield experiment. The listeners had symmetric sensorineural hearing loss and were experienced users of bilateral hearing aids. The task was to identify the speech of a target talker located directly in front of the listener (0 deg azimuth) in the presence of two masker talkers co-located with the target or spatially separated at ±90 deg azimuth. In this symmetric spatial configuration of sound sources, there was not an overall difference in target-to-masker ratio (T/M) at either ear during stimulus presentation, thus eliminating any "better ear" advantage. Reverberation was varied by physically changing the sound absorption characteristics of the sound booth in which the experiments were conducted. Four listening conditions were tested: unaided, bilateral aided, unilateral aided-right, and unilateral aided-left. In general, the hearing aid users demonstrated less spatial release than normal-hearing listeners although there were large inter-subject differences across the different aided conditions. [Work supported by NIH/NIDCD and AFOSR.]

1:50

3pPP4. Verification of the assumption of frequency-independent recovery from forward masking. Magdalena Wojtczak and Andrew J. Oxenham (Dept. of Psych., Univ. of Minnesota, Elliott Hall N218, 75 East River Rd., Minneapolis, MN 55455)

Psychophysical methods for deriving basilar-membrane (BM) input-output functions from forward-masking data assume that the recovery from forward masking at a given probe-frequency place is independent of the masker frequency. The validity of this assumption was tested here. The levels of an on-frequency (4-kHz) and off-frequency (2.4-kHz) forward masker were individually adjusted to yield the same masked threshold for the probe presented immediately after the masker offset. Once the masker levels had been established, probe thresholds were measured using these maskers for masker-probe delays between 0 and 115 ms. To prevent off-frequency listening, a low-level notched noise centered on the probe frequency was presented throughout each trial. Possible confusion effects were reduced by presenting a contralateral 7-kHz tone synchronous with the masker. Since the recovery functions started with equal probe levels, any differences between the recovery functions for the on- and off-frequency masker would imply masker-frequency or masker-level dependence of forward-masking recovery. Preliminary data suggest some listeners may exhibit a slower recovery for the high-level (92-dB SPL) off-frequency masker than for the lower-level on-frequency masker. Implications for BM compression estimates derived from psychophysical forward masking will be discussed. [Work supported by NIH grant R01DC03909.]

2:05

3pPP5. Vision influences judgments of acoustic duration. Michael Schutz and Michael Kubovy (Univ. of Virginia, 102 Gilmer Hall, Charlottesville, VA 22904)

We video-recorded isolated marimba notes performed by a world-renowned percussionist attempting to produce long (L) and short (S) notes. We generated audio-visual stimuli by crossing the video components of these notes. Subjects rated the duration of sounds presented in the audio-alone (A), and in the audio-visual (AV) condition (in the latter, they were to rely only on auditory information). In the A condition the S and L sounds were rated to be equal. In the AV condition, visual information affected ratings, but auditory information did not. In a second experiment, to undermine a response-bias account of these findings, we introduced a temporal offset between the auditory and visual information. When the visual information preceded the auditory, vision influenced auditory judgments, but when the visual information lagged it did not. This tolerance for visual lead may reflect an ecology in which optical information about an event precedes the acoustic information about it. Together, these studies demonstrate a novel visual influence on auditory judgments of duration through the integration of sensory information.

2:20

3pPP6. Detection of signals with expected and unexpected total durations and temporal structures. Beverly A. Wright (Dept. of Commun. Sci. and Disord. and Inst. for Neurosci., 2240 Campus Dr., Northwestern Univ., Evanston, IL 60208-3550, b-wright@northwestern.edu) and Huanping Dai (Univ. of Wyoming, Laramie, WY 82071)

When trying to detect a tonal signal in a continuous broadband noise, listeners attend selectively to the total duration of the expected signal, but do they also attend selectively to its temporal structure? To investigate this question, a probe-signal method was used to measure the detectability of expected 1-kHz target signals that differed across conditions in total duration (4 or 299 ms) and temporal structure (pure tone or pulse train), and unexpected 1-kHz probe signals that differed from the target in total duration, temporal structure, or both. For all targets, probe detectability de-

creased as the difference between the target and probe duration increased. However, the pattern of results differed between the two target durations. For the 299-ms targets, probe detectability did not differ significantly between pure-tone and pulse-train probes, regardless of the target type or the probe duration. In contrast, for the 4-ms targets, detectability for each pulse-train probe was significantly higher than for pure-tone probes of the same duration. Thus, listeners appear to attend to the targets total duration more than its temporal structure when the target is long, but do the reverse when the target is short. These results are inconsistent with current temporal-integration models. [Work supported by NIH.]

2:35

3pPP7. Temporal integration and compression near absolute threshold in normal and impaired ears. Christopher J. Plack (Dept. of Psych., Lancaster Univ., Lancaster, LA1 4YF, England, c.plack@lancaster.ac.uk) and Vicki Skeels (The Hearing Care Ctr., Colchester, CO1 1NU, England)

The decrease in absolute threshold with increasing stimulus duration is greater for listeners with normal hearing than for listeners with sensorineural hearing loss. It has been suggested that the difference is related to reduced basilar-membrane (BM) compression in the impaired group. This hypothesis was tested by comparing temporal integration and BM compression near absolute threshold in normal and impaired ears. In the first phase, absolute thresholds were measured for 4-, 24-, and 44-ms pure-tone signals, with frequencies (fs) of 2 and 4 kHz. The difference between the absolute thresholds for the 4- and 24-ms signals was used as a measure of temporal integration. In the second phase, compression near threshold was estimated by measuring the level of a 200-ms off-frequency (0.45 fs) pure-tone forward masker required to mask a 44-ms pure-tone signal presented at sensation levels of 5 and 10 dB. There was a significant negative correlation between the amount of temporal integration and absolute threshold, but there was no correlation between absolute threshold and compression at low levels; both normal and impaired ears showed a nearly linear response. The results suggest that the differences in integration between normal and impaired ears cannot be explained by differences in BM compression.

2:50

3pPP8. Development of temporal-interval discrimination during adolescence. Julia Jones Huyck (Dept. of Commun. Sci. and Disord., 2240 Campus Dr., Northwestern Univ., Evanston, IL 60208-3550, j-huyck@northwestern.edu) and Beverly A. Wright (Dept. of Commun. Sci. and Disord. and Inst. for Neurosci., Northwestern Univ., Evanston, IL 60208-3550)

The accurate perception of time intervals in the 10–100-ms range plays an important role in sensory encoding, including the processing of speech, yet little is known about the development of this ability, particularly during adolescence. Here, auditory temporal-interval discrimination thresholds were measured in normally developing individuals aged 11 (n = 14), 14 ($n = 8$), and 18–26 ($n = 17$) years. Both average performance (mean within-listener threshold) and performance consistency (within-listener standard deviation) were evaluated for each listener with a 100-ms standard at each of two frequencies, 1 and 4 kHz. At the group level, for both frequencies, average performance improved significantly, from ~ 48 to ~ 22 ms, between 11 and 14 years of age (both $p \leq 0.02$), but did not improve further between 14 years and adulthood (both $p \geq 0.92$). There were parallel improvements in performance consistency. At the individual level, while the majority of 11-year-olds showed poorer performance than 14-year-olds and adults on both measures, a sizable minority performed within the range of adults. Thus it appears that, both in terms of average performance and performance consistency, the sensitivity to differences in temporal intervals has a developmental time course that can differ considerably across individuals, but, for most, reaches well into adolescence. [Work supported by NIH/NIDCD.]

Session 3pSC

Speech Communication: Atypical Populations; Acoustic and Physiologic Modeling (Poster Session)

Emily Q. Wang, Cochair

Rush Univ., Communication Disorders and Science, 1653 W. Congress Parkway, Chicago, IL 60612

Julie M. Liss, Cochair

Arizona State Univ., Dept. Speech and Hearing Science, Box 871020, Tempe, AZ 85287-1020

Contributed Papers

All posters will be on display from 1:00 p.m. to 3:20 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 2:10 p.m. and contributors of even-numbered papers will be at their posters from 2:10 p.m. to 3:20 p.m.

3pSC1. Phonetic context effects in normal-hearing listeners using acoustic simulations of cochlear implant signal. Radhika Aravamudhan (PCO School of Audiol., 8360 Old York Rd., Elkiins Park, PA 19454, raravamudhan@pco.edu) and Andrew J. Lotto (Univ. of Arizona, Tucson AZ, 85721-0071)

Recent work has indicated that the perception of speech sounds shifts as a function of preceding and following context and that these shifts are due in part to the particular spectral makeup of the context sounds [Holt *et al.*, J. Acoust. Soc. Am. **108**, 710–722 (2000)]. Aravamudhan and Lotto [J. Acoust. Soc. Am. **118**, 1962–1963 (2005)] studied these context effects in listeners with cochlear implants (CI) and the results demonstrated absent or abnormal context-dependent speech perception. The lack of normal context effects in CI users may have practical implications for situations in which there is substantial coarticulation (e.g., nonlaboratory speech) or talker variability (e.g., switching between multiple speakers). Since the representation of speech in CI listeners differs from NH listeners, the current study was designed to investigate the effects of processing normal signals through a CI processor with NH listeners. In particular, we investigated how the spectral resolution of implant input could limit context-dependent perception even in NH listeners. The results showed that the context effects elicited in NH listeners using CI processed signals were similar to CI listeners. The effects of context were moderated by the processing strategy being simulated as well as some of the parameter settings.

3pSC2. The role of prosody in the foreign accent syndrome: A single-subject study. Diane M. Garst, June S. Levitt, and William F. Katz (Univ. of Texas at Dallas, Callier Ctr. for Commun. Disord., 1966 Inwood Rd., Dallas, TX 75235, wkatz@utdallas.edu)

Foreign accent syndrome (FAS) is an unusual motor speech disorder involving a perceived foreign accent resulting from brain damage. An important question concerning FAS is the role of prosodic deficits, including the extent to which difficulties in stress assignment may contribute to segmental errors. In this study, acoustic analyses were performed on the speech of a 46-year-old female patient who presented with a pseudo-accent described as Swedish. Stop consonant VOT, consonant burst spectra and duration, and vowel formant frequencies and trajectories were analyzed, along with prosodic cues for lexical stress assignment and sentence-level intonation. Results indicated that the patients vowels were centralized and realized with reduced dynamic specifications. There was a strong tendency to realize the English alveolar flap as a full stop, and to produce flaps that had greater-than-normal closure durations. Lexical stress assignment was frequently inaccurate and highly variable, with

similar problems noted for non-speech stimuli. Overall, the data suggest that stress assignment deficits were the basis for most of her segmental difficulties. The findings support the view that local prosodic deficits play a critical role in FAS.

3pSC3. Effects of accent-reduction techniques for the treatment of an individual with foreign accent syndrome. William F. Katz, Diane M. Garst, Karen E. Kaplan, and Emily J. Frisch (Univ. of Texas at Dallas, Callier Ctr. for Commun. Disord., 1966 Inwood Rd., Dallas, TX 75235, wkatz@utdallas.edu)

Foreign accent syndrome (FAS) is a rare speech motor disorder characterized by the emergence of a perceived foreign accent following brain damage. Although FAS has received increased attention in recent years, the issue of intervention for this disorder has not been systematically explored. In this study, an English-speaking female patient with FAS thought to sound Swedish was given short-term accent reduction treatment for selected front vowels (/i/, /ɪ/, /ɛ/), back vowels (/u/, /ʊ/, /o/), and interdental consonants (/θ/, /ð/). Therapy consisted of structured practice on word lists containing targeted phonemes. A list of compound nouns was included as untreated control items. A multiple-baseline across behaviors design was used to determine treatment efficacy. Probes were included for response generalization, maintenance, and long-term retention. Digitally-recorded productions by the FAS talker were played in randomized order to a phonetically-trained listener who rated a targeted sound in each word as foreign-sounding or not. Results indicated highly variable performance on treated items, generalization probes, and control probes throughout the course of the experiment, with no apparent acquisition or maintenance resulting from the intervention. These findings suggest that structured intervention techniques were not successful for this patient.

3pSC4. Quantifying speech rhythm deficits in dysarthria. Julie Liss, Stephanie Spitzer, Kaitlin Lansford, Yu-kyong Choe, Kathryn Kennerley (Motor Speech Disord. Lab., Arizona State Univ., Box 870102, Tempe, AZ 85287-0102, julie.liss@asu.edu), Sven Mattys, Laurence White (Univ. of Bristol, Bristol BS8 1TN, UK), and John Caviness (Mayo Clinic Arizona, Scottsdale, AZ 85259)

Disturbances in speech rhythm are common in dysarthria, and deficit patterns vary among dysarthria subtypes. Attempts to quantify differences in disordered rhythm patterns have met with limited success. Here, we attempt to differentiate various forms of dysarthria using metrics designed to quantify linguistic rhythm on a stress-timed versus syllable-timed continuum. Speakers with hypokinetic, ataxic, hyperkinetic, or mixed flaccid-

3134   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3134**

spastic dysarthria read sentences. Rhythm metrics were calculated for these utterances based on vocalic and consonantal interval durations. Differences among groups were discovered for %V (percentage of utterance duration comprised of vocalic intervals), VarcoV (rate-normalized standard deviation of vocalic interval durations), and nPVI-V (normalized pairwise variability index for vocalic intervals). Differences were independent of variable speaking rates among groups. All dysarthric groups produced rhythm patterns less stress-timed than the control American English speakers, with lower VarcoV and nPVI-V and higher %V. Those with flaccid-spastic dysarthria secondary to amyotrophic lateral sclerosis exhibited a syllable-timed pattern more extreme than even control Spanish speakers, indicating minimal differentiation of stressed and unstressed syllables. These results show that the various types of rhythmic disturbance accompanying dysarthria can be effectively described and distinguished within a stress-/syllable-timed continuum and relative to healthy speech.

3pSC5. An improved distance metric for disordered speech evaluation. Savyasachi Singh, John Harris (CNEL, Dept. of Elec. and Comput. Eng., Univ. of Florida, Gainesville, FL 32611), Rahul Shrivastav, and Christine Sapienza (Univ. of Florida, Gainesville, FL 32611)

Assessment of the quality of disordered speech due to dysarthria, laryngeal diseases, stroke or other impairments is important for feedback to keep track of the development of patients pathology. Methods based on perceptual evaluation of speech quality are accurate and robust but expensive and time consuming. This study focuses on the development of a distance metric for disordered speech quality assessment based on signal processing techniques which is inexpensive, fast, affordable and easy to use. Previous efforts from our group have shown this can be achieved with short-term Mel frequency cepstral coefficients (MFCC) along with dynamic time warping (DTW) and Itakura-Saito distortion measures. Disordered speech is characterized by several long time signal variations such as increase of hoarseness in voice, speaking rate variability, word closure variability, presence of high frequency noise components, large pitch period and peak amplitude variations. An improved distortion metric is developed which accounts for these characteristics by augmenting the feature set with measures such as jitter, shimmer, amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ), harmonics to noise ratio (HNR) and uses DTW for alignment. Initial results with a speech verification database are promising while the full system development is underway.

3pSC6. Consonant-vowel (CV) and vowel-consonant (VC) transitions in the speech of persons with dysarthria and healthy controls. Gary Weismer, Christina Kuo, and Yun-Ching Chung (Dept. of Communicative Disord. and Waisman Ctr., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706)

Previous work has shown that speakers with dysarthria have shallower-than-normal slopes for second formant transitions in diphthongs and semivowels. Little is known, however, about transition characteristics in dysarthria for CV and VC transitions. In the present work, stop-vowel, fricative-vowel, and nasal-vowel formant transitions (and VC counterparts) were measured from words in a reading passage produced by 25 controls, 22 patients with Parkinson disease, and 15 patients with ALS. Measures were made over a 50-ms interval extending from the first glottal pulse of the vowel (in CV syllables) or back from the final glottal pulse in VC syllables. Transition extents over this interval were measured and slopes derived. Distributional analyses suggest that slope differences between control speakers and speakers with dysarthria are subtle, but show a slight tendency for greater absolute values for the control speakers. This is consistent in kind with work on diphthongs and semivowels, but the subtlety of the effect is quite different. The apparent greater resistance of the CV and VC transitions to the effects of dysarthria, compared to effects on other kinds of transitional events, is discussed within the framework of contemporary speech production theory. [Work supported by NIH Award DC 03723.]

3pSC7. Evaluating the effect of subthalamic deep brain stimulation on vocal function in Parkinson's disease. Emily Wang (Rush Univ. Medical Ctr., Chicago, IL 60612), Jack Jiang (Madison, WI 53706), Leo Verhagen Metman, Roy Bakay, and Bryan Bernard (Rush Univ. Medical Ctr., Chicago, IL 60612)

The study reports the effect of subthalamic deep brain stimulation (STN DBS) on vocal function in Parkinson's disease (PD). Specifically, the effect was evaluated in relation to the vibratory characteristics of the vocal folds during sustained vowel phonation using two different methods: one measures jitter and shimmer, while the other measures the correlation dimension D2, a nonlinear dynamic analysis. Twenty right-handed individuals with advanced PD and dysarthria underwent unilateral STN DBS. Ten were operated on the right hemisphere and ten on the left hemisphere. The side receiving STN DBS had more affected motor function. Speech was evaluated before surgery and three-to-six months after surgery with stimulator-off and with stimulator-on; all were off anti-Parkinsonian medication for 12 h before evaluation. Evaluators and subjects were blinded to the subjects stimulator status at the postsurgery evaluations. Nonspeech motor performance was assessed with UPDRS-III. Each subject produced six maximally sustained "ah" phonations (MSVPs) at his/her habitual loudness level. Two-second samples extracted from the MSVPs were used for both analyses. In addition, the maximum vocal intensity, duration, mean F0, and cycle-to-cycle F0 variation were also obtained. The results were discussed with regard to Chaos theory and the phonation characteristics in PD.

3pSC8. Word recognition in dysarthric speech: Evidence for the time-course hypothesis. Kaitlin Lansford, Julie Liss, Stephanie Spitzer (Motor Speech Disord. Lab., Arizona State Univ., Box 870102, Tempe, AZ 85287-0102, julie.liss@asu.edu), Sven Mattys (Univ. of Bristol, Bristol BS8 1TN, UK), and John Caviness (Mayo Clinic Arizona, Scottsdale, AZ 85259)

The detrimental effect of speaker variability on word recognition has been reported using a wide variety of tasks and stimuli. In this study, the use of dysarthric speech alongside unimpaired speech allowed us to examine the effect of indexical variability from a new perspective. We compared voice-specificity effects in words produced by healthy speakers and the same words produced by dysarthric patients suffering from Huntington's disease (HD). The speech of persons with HD is particularly suited to this issue because of its idiosyncratic and variable manifestation. Reaction time and accuracy results provided confirmatory evidence for Luce *et al.*'s (2003) time-course hypothesis: The slower responses occasioned by processing dysarthric stimuli were accompanied with greater reliance on surface details of the stimuli. Thus, specificity effects can be seen as resulting from a coping strategy called upon when the system is presented with an input that is highly dissimilar from canonical lexical representations. In addition, we found some evidence that the time-course hypothesis holds even when intelligibility is controlled. This finding confirms that time, rather than processing effort, is the cause of specificity effects, with activation confined to abstract representations in an early stage and spreading to episodic traces later on.

3pSC9. Role of glottal-pulse rate, vocal-tract length, and original talker upon judgements of speaker sex and age. David R. R. Smith, Thomas C. Walters, and Roy D. Patterson (CNBH, Physiol., Development & Neurosci., Univ. of Cambridge, Downing St., Cambridge, CB2 3EG, UK, drrs2@cam.ac.uk)

Glottal-pulse rate (GPR) and vocal-tract length (VTL) are important determinants of the perceived sex and age of the speaker [D. R. R. Smith and R. D. Patterson, J. Acoust. Soc. Am. **118**, 3177–3186 (2005)]. Our previous research simulated the voices of variously-sized speakers of both sexes by manipulating the recorded vowels of one adult male talker. The current study explored whether there are additional cues in the voices of men, women, and children that influence judgements of speaker sex and age. We manipulated the recorded vowels of an adult man, adult woman,

3p WED. PM

young boy, and young girl, and determined the effect upon the distribution of sex and age responses (man, woman, boy, girl). Results show that the distribution of sex and age judgements across the GPR-VTL plane is heavily influenced by GPR and VTL, but it is also affected by the original talker's size (or age). The effect of original talker appears to be mainly due to the consistent difference between oral-pharyngeal length ratios of children and adults. We also report gender differences in male and female listeners across the GPR-VTL plane; adult men and women listeners show somewhat different biases. [Research supported by UK MRC (G9900369, G0500221) and German Volkswagen Foundation (1/79 783).]

3pSC10. Modeling the effects of frequency shifts on vowel identification. Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada) and Peter F. Assmann (Univ. of Texas at Dallas, Box 830688 Richardson, TX 75083)

Previous experiments examining the effects of frequency shifts on vowel perception show that identification accuracy drops when the spectrum envelope is shifted up by more than about 150%, or shifted down by factors smaller than about 60% relative to adult male ranges. Such shifts produce formant patterns near the extreme limits found in human voices. But these effects interact with fundamental frequency (F0): in some conditions identification accuracy is improved by shifting the formant frequencies (FFs) and F0 in the same direction, compared to conditions where one is raised and the other is lowered. The results indicate the presence of perceptual mechanisms that are sensitive to the natural covariation of F0 and FFs in human voices. Initial modeling shows that including F0 and FFs predicts listeners' behavior better than FFs alone. Specifically, posterior probabilities from linear discriminant function analysis are better correlated with listeners' identification rates when F0 is included than when it is not. Further modeling suggests prediction of overall perceptual results generally improves (especially in mismatched conditions) for modified models that include a positive correlation between F0 and FF that is somewhat weaker than that observed in natural speech databases. [Work supported by NSF and SSHRC].

3pSC11. Laryngeal airflow modeling; an investigation based on computational aeroacoustics. Mihai Mihaescu, Ephraim Gutmark (Dept. of Aerosp. Eng., Univ. of Cincinnati, 310-799 Rhodes Hall, P.O. Box 210070, Cincinnati, OH 45221), Sid Khosla (Univ. of Cincinnati, Cincinnati, OH 45267), and Ronald Scherer (Bowling Green State Univ., Bowling Green, OH-43403)

The paper presents an efficient method for computational aeroacoustics applied to simulate the flow and the acoustics for a static laryngeal model considering the vocal folds widely open. The work analyzes the whisperlike process defined as the sound generated by the turbulent glottal airflow in the absence of vocal fold vibration. A decomposition of the flow variables is used that allows separation of flow and acoustic computations. The approach consists in solving the low Mach number flow field by incompressible large eddy simulation. This provides the possibility to compute the acoustic sources. The inhomogeneous wave equation derived from the fundamental equations of motion for a compressible fluid is employed to compute the acoustic field. The purpose of the study is to provide realistic numerical predictions of the flow patterns and the generated acoustic field in the glottis and the vocal tract regions.

3pSC12. The nonlinear relation between articulation and acoustics: Multiple constrictions. Khalil Iskarous (Haskins Labs., 300 George St., New Haven, CT 06511)

The vocal tract can be usefully approximated as a linear filter, but it has been known since the advent of this approximation that the relation between the area function and the acoustic output is nonlinear, despite linearity with respect to the input. This nonlinearity can be readily seen in nomograms displaying formants changing as a contriction changes its lo-

cation or degree. It will be shown in this work that the nonlinear relation between area function and acoustics becomes even more significant when several constrictions are superimposed, which is the norm in natural speech production, where constrictions overlap in time. Simulations have been carried out, using a Kelly-Lochbaum digital lumped-loss transmission line, that show a complex relation between the formants resulting from a multiply constricted vocal tract to formants due to the constrictions when each is present by itself. The goal of the simulations is to predict time varying formant patterns when 2 or 3 constrictions are overlapped and all changing in time. A theory is then presented for how time-varying constrictions add in the acoustic domain. Implications for restrictions on compensation in speech production will be discussed. [Work supported by NIH NIDCD 02717.]

3pSC13. New features in VTAR: A Matlab-based computer program for vocal tract acoustic modeling. Carol Espy-Wilson, Xinhui Zhou (Dept. of Elec. and Computer Eng., Univ. of Maryland, College Park, MD 20742, espy@glue.umd.edu), Mark Tiede (Haskins Labs, New Haven, CT 06511), and Suzanne Boyce (Univ. of Cincinnati, Cincinnati, OH 45202)

VTAR, a Matlab-based computer program for vocal tract acoustic response calculation [Zhou *et al.*, J. Acoust. Soc. Am. **115**(5), 2543 (2004)], has been used in both research and teaching. In its latest version, several new features are included, which are speech sound synthesis with source model options, formant sensitivity function calculation, susceptance plots calculation useful particularly for nasalized vowel analysis, and addition of a new set of area function data for liquid sounds extracted from MR (magnetic resonance) images. These new features along with the user-friendly interface significantly enhance the usability of VTAR for both teaching and research purposes. [Work supported by NIH Grant 1 R01 DC05250-01.]

3pSC14. Structural analysis of a three-layered vocal fold model. Douglas Cook, Eric Nauman, and Luc Mongeau (Ray W. Herrick Labs., Purdue Univ., 140 S. Intramural Ave., W. Lafayette, IN 47907-2031)

Although many detailed fluid-structure interaction models of human phonation have been created, structural analysis of the vocal fold structure is uncommon. As a result, the vocal fold structure is not well understood. Structural analysis provides useful information for the creation of efficient fluid-structure interaction and reduced order models. To this end, a detailed model of the human vocal fold was created, and modal analysis was used to determine the model's sensitivity to boundary conditions, geometric parameters, and material properties.

3pSC15. Intraglottal pressures in a static physical hemilarynx model. Ken J. DeWitt, Yang Bo (Dept. of Chemical Eng., Univ. of Toledo, Toledo, OH 43506), Pushkal Thapa, Lewis P. Fulcher, and Ronald C. Scherer (Bowling Green State Univ., Bowling Green, OH 43403)

Intraglottal air pressures during oscillation of a single vocal fold against an immobile opposing structure are of interest to surgery after unilateral vocal fold removal and the validity of pressure data in hemilaryngeal bench studies. Intraglottal pressures and flows within a static hemilarynx were studied using Plexiglas model M5 with removable vocal folds and FLUENT, a commercial code using the control-volume technique. For a minimal diameter of 0.04 cm, an axial glottal length of 0.3 cm, and a vertical static opposing wall, data were collected for converging and diverging angles of 5, 10, and 20 (and 0) deg. Pressure distributions were obtained for transglottal pressures of 3 to 40 cm water. For divergent, zero, and smaller converging angles within the physical model, greater intraglottal velocities could be coaxed to favor one side or the other and were typically stable. The side of faster flow for divergent angles had lower pressure in the downstream third of the glottis ranging from ap-

proximately 5% to 20% of the transglottal pressure, which would produce asymmetrical forces acting on the opposing vocal folds. FLUENT results are in good agreement with the measured pressures. [Work supported by NIH R01DC03577.]

tence used for the training model. Finally, when there was a mismatch between whispered speech being used in training and normally phonated speech in testing, the SID performance was severely impacted.

3pSC16. External factors impacting the performance of speaker identification: Multisession audio research project (MARP) corpus experiments. Stanley Wenndt (Air Force Res. Lab., 525 Brooks Rd., Rome, NY 13441), Aaron Lawson, and Allen Stauffer (RADC, Rome, NY 13441)

This study looks at the effect of data conditions on automated speech processing systems. The goal is to better understand the impact of acoustical features on accuracy and to develop more robust features. A speaker identification (SID) system was used for the experiments. To explore this issue, a new longitudinal database was collected involving 60 speakers over 18 months. This corpus allowed us to examine four data factors that impact SID: (1) intersession variability, (2) question intonation (3), text-dependency (identical phonetic content), and (4) whispered speech. First, we found that intersession SID suffered an average loss in accuracy of 17% independent of the time latency between sessions. Second, differing the intonation conditions in train and test hurt SID performance by 5%. Third, text-dependent data showed the most dramatic impact, where using phonetically-identical test and train sentences yielded 0% error. However, replacing the target speakers content with a random text-independent sentence, *ceteris paribus*, caused accuracy to plummet 94%. In all of the erroneous identifications, the top-ranked speaker was speaking the sen-

3pSC17. Aerodynamics at glottal exit using particle image velocimetry method in a canine larynx. Sid Khosla, Shanmugam Murugappan, R. R. Lacham Raju, and Ephraim J. Gutmark (Dept. of Otolaryngol., Univ. of Cincinnati Medical Ctr., Cincinnati, OH 45267)

There are several excellent computational and mechanical models that describe the flow fields exiting the glottis during phonation. However, these models need to be validated and possibly refined in the canine larynx. Experimental characterization of flow at the glottal exit has been limited to hot-wire anemometry measurements made no closer to the glottal exit than 1 cm; these measurements have also been limited to magnitude and have not included flow direction. Using the particle image velocimetry method, velocity fields have successfully been measured in the excised canine larynx during phonation within the first cm above the glottal exit. Previous papers have discussed the methodology and qualitatively described the spatial and temporal location of the vortices. In this paper, quantitative characteristics of the velocity field in the first centimeter above the glottal exit in an excised canine larynx model without a vocal tract will be shown; these characteristics elucidate how the jet transforms as a function of space and phase. Comparisons to physical and computational models will also be described and further work will be discussed. [Work supported by NIH Grants K08DCC005421 and R01DC003577.]

3p WED. PM

Plenary Session, Business Meeting, and Awards Ceremony

Anthony A. Atchley, President
Acoustical Society of America

Annual Business Meeting of the Acoustical Society of America

Presentation of Certificates to New Fellows

Michael R. Bailey	Hedwig E. Gockel
Paul E. Barbone	Kai Ming Li
Damian J. Doria	T. Douglas Mast
John M. Eargle	Albert Migliori
Stephen J. Elliott	Claire Prada
	Karim Sabra

Presentation of Acoustical Society Awards

Medwin Prize in Acoustical Oceanography to Brian S. Dushaw

Rossing Prize in Acoustics Education to William J. Strong

R. Bruce Lindsay Award to Dorian S. Houser

Helmholtz-Rayleigh Interdisciplinary Silver Medal to Edwin L. Carstensen

Gold Medal to Katherine S. Harris

Presentation of Vice President's Gavel

Presentation of President's Tuning Fork

ACOUSTICAL SOCIETY OF AMERICA

R. BRUCE LINDSAY AWARD
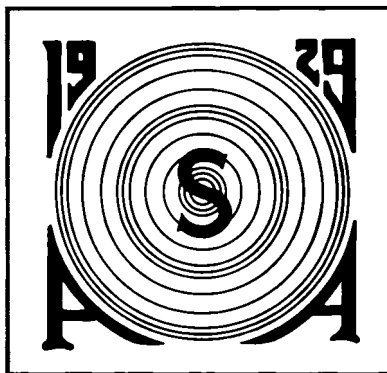
Dorian S. Houser

2007

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	William E. Cooper	1986
Leo L. Beranek	1944	Ilene J. Busch-Vishniac	1987
Vincent Salmon	1946	Gilles A. Daigle	1988
Isadore Rudnick	1948	Mark F. Hamilton	1989
J. C. R. Licklider	1950	Thomas J. Hofler	1990
Osman K. Mawardi	1952	Yves H. Berthelot	1991
Uno Ingard	1954	Joseph M. Cuschieri	1991
Ernest Yeager	1956	Anthony A. Atchley	1992
Ira J. Hirsh	1956	Michael D. Collins	1993
Bruce P. Bogert	1958	Robert P. Carlyon	1994
Ira Dyer	1960	Beverly A. Wright	1995
Alan Powell	1962	Victor W. Sparrow	1996
Tony F. W. Embleton	1964	D. Keith Wilson	1997
David M. Green	1966	Robert L. Clark	1998
Emmanuel P. Papadakis	1968	Paul E. Barbone	1999
Logan E. Hargrove	1970	Robin O. Cleveland	2000
Robert D. Finch	1972	Andrew J. Oxenham	2001
Lawrence R. Rabiner	1974	James J. Finneran	2002
Robert E. Apfel	1976	Thomas J. Royston	2002
Henry E. Bass	1978	Dani Byrd	2003
Peter H. Rogers	1980	Michael R. Bailey	2004
Ralph N. Baer	1982	Lily M. Wang	2005
Peter N. Mikhalevsky	1984	Purnima Ratilal	2006

CITATION FOR DORIAN S. HOUSER

. . .for contributions to animal bioacoustics and to understanding echolocation and hearing by dolphins.

SALT LAKE CITY, UTAH • 6 JUNE 2007

Dorian Scott Houser completed his B.A. in Biology (Summa Cum Laude) at Coker College in June 1992 and his PhD at the University of California, Santa Cruz (UCSC) in 1998. During his graduate tenure at UCSC Dorian received a fellowship from the American Society for Engineering Education that allowed him to work during summers at his choice of naval research laboratories. Dorian chose to work at the Navy Marine Mammal Program at the Space and Naval Warfare Systems Center, San Diego (SSC-SD), where he hoped to pursue his interest in marine mammal diving physiology. He sought out Sam Ridgway, an internationally known and respected expert in marine mammal physiology, to see if there was an opportunity to work during his first summer under the fellowship studying diving pinnipeds. Unfortunately, Sam had no ongoing diving physiology tasks so he pointed Dorian to the Biosonar Program that had just begun a few years before at the laboratory.

In that first summer Dorian began looking at the emitted echolocation signals of several dolphins. A paper had just been published that segregated false killer whale echolocation clicks into three categories and Dorian was determined to see if the application of this scheme to dolphin echolocation could be used to look for systematic variations in echolocation click production. Dorian attacked this project with enormous energy (an approach he continues to apply to his projects) and began an exhaustive study into the basics of dolphin bioacoustics and signal processing. His efforts were rewarded by his first publication in the *Journal of the Acoustical Society of America* (JASA) in 1999.

After obtaining his Ph.D., Dorian was awarded a National Research Council Postdoctoral Research Fellowship to continue bioacoustics research at SSC-SD and since then, Dr. Houser's contributions to the Society and to animal bioacoustics have been steady with approximately one quarter of his published works appearing in the Journal. His research has moved to the leading edge of our understanding of the effects of sound on marine mammals. Dorian developed a basis for modeling hearing in whales by using multi-threaded, evolutionary computational methods to model the hearing system of the dolphin. From that point, he ground-truthed the model's goodness-of-fit to the dolphin and applied it to the humpback whale, a species for which only anatomical data existed. For his work Dorian was acknowledged as ''key personnel'' in the Strategic Environmental Research and Development Program (SERDP) ''Project of the Year'' in 2000.

After completing his postdoctoral fellowship, Dorian formed his own company, Biomimetica, and has continued to be an integral part of the Navy's exploration into new areas of psychoacoustic measures of the hearing of marine mammals. Dorian, along with Sam Ridgway, applied medical imaging techniques to explore the physiology and anatomy of the dolphin signal generating and receiving systems. These studies marked the first time structural (MRI, CT) and functional brain scans (PET, SPECT) were performed on a living cetacean. The results had far-reaching implications, ranging from highlighting differences between post mortem and *in vivo* anatomy, describing *in vivo* the complex nasal sacs around the ears and their relationship to directional hearing, to hypothesizing on the possibility of thermoregulation of acoustic lipids in the dolphin's melon and its implications for sound transmission.

Dorian also helped develop and implement the ''Biosonar Measurement Tool,'' a device that captured the complete acoustic environment (signal emissions and echo reception) of a free-swimming dolphin in open-water target detection tasks. It also recorded the three-dimensional motion of the animal through the water enabling a virtual recreation of the animal's search and bioacoustic behavior. The results, published in JASA, showed how echolocating dolphins employed their biosonar during open-ocean target detection and provided insight for biologically inspired search and detection algorithms used in newly developed biomimetic sonar.

Dorian also applied his expertise in physiology to advance electrophysiological methods for hearing assessment. He is co-investigator on an effort to develop and standardize the measurement of auditory evoked potentials from marine mammals. These techniques have led to the first large scale, controlled study of marine mammal audition, with hearing tests to date being conducted on

more than 42 dolphins. The results of this study revealed important relationships between animals' hearing abilities and age, gender, ancestry, and medical history and have provided key inventory management data for the Navy Marine Mammal Program.

Dorian is active in investigating potential physiological causes of marine mammal strandings. His 2001 paper in the *Journal of Theoretical Biology* has served as a catalyst for hypotheses explaining beaked whale strandings involving acoustically driven or activated bubble growth. Dorian is currently investigating the potential for this phenomenon by attempting to ultrasonically measure intravascular nitrogen bubbles that may form in a diving animal just after it surfaces from a series of dives. The results of this project should provide some basic answers to questions regarding nitrogen accumulation and bubble growth in diving marine mammals and lay the groundwork for understanding the potential impact of sonar on beaked whales.

In addition to possessing excellent research and experimental skills, Dr. Houser also has the important ability to maintain focus on the larger context of the research, apply the results to real-world problems, and publish work in publicly accessible, peer-reviewed literature. He has been a key participant in the development of several Navy Environmental Impact Statements assessing the potential effects of acoustic activities on marine mammals. In these technically challenging and often emotionally charged efforts, Dorian has shown tremendous skill in explaining to a lay audience complex acoustic topics ranging from acoustic metrics to the process of rectified diffusion, all while maintaining a positive and professional relationship with Navy environmental planners, government regulators, and non-governmental environmental groups.

In summary, Dr. Houser's work in marine mammal bioacoustics has made a significant impact on the state of knowledge regarding marine mammal audition and the physiological effects of sound on diving marine mammals. Dorian is active in the Society as a presenter and member of the Animal Bioacoustics Technical Committee. We believe he is an outstanding researcher with unparalleled ability and is a worthy recipient of this R. Bruce Lindsay Award.

PATRICK W. MOORE
JAMES J. FINNERAN

ACOUSTICAL SOCIETY OF AMERICA

Helmholtz-Rayleigh Interdisciplinary

Silver Medal in

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics

Edwin L. Carstensen

2007

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 EDWIN L. CARSTENSEN

. . . for contributions to the physics of biomedical ultrasound.

SALT LAKE CITY, UTAH • 6 JUNE 2007

Edwin Lorenz Carstensen was born on December 8, 1919, to Opal and August Carstensen in Oakdale, Nebraska. In his small hardware and appliance store, Ed's father sold and repaired radios; that may have started Ed's interest in science and math. Also interested in music, he studied clarinet, piano, saxophone, and voice.

Intent on becoming a high school science and music teacher, Ed attended Nebraska State Teachers' College in Wayne, Nebraska. A five-minute conversation with his physics professor, who encouraged him to attend graduate school to study the physics of music, proved to be a turning point in Ed's career. After earning his B.S. in 1941, he headed to Case School of Applied Science (Cleveland) to study physics. He took an acoustics course from Robert Shankland, but about this time World War II intervened. Dr. Shankland was recruited to form the Underwater Sound Reference Laboratories (USRL) for the U.S. Navy, in Orlando, Florida. He recruited Ed to come with him, and soon Ed became Head of the Laboratory Section of USRL. The acoustical properties of bubbles in submarine wakes was a prime concern, and out of the work came the classic Carstensen-Foldy paper on sound propagation through bubbly water (JASA, 1947). Thus began Ed's interest in the interaction of sound and bubbles. While at USRL, he completed his M.S. in Physics, and also met and married (August 1, 1947) his wonderful wife, Pam McDonald Carstensen.

In 1948, Ed went to the University of Pennsylvania, attracted by a collaborative graduate program between the Departments of Physical Medicine and Electrical Engineering. There he met his mentor Herman Schwan and began his career-long interest in the biological effects of ultrasound and electromagnetic fields. After Ed and Dr. Schwan first compared ultrasonic and dielectric methods of diathermy, Ed began his dissertation research on the acoustic properties of blood. He discovered that ultrasound absorption occurs at the macromolecular level. His groundbreaking work pioneered our understanding of relaxation mechanisms in the absorption of ultrasound in blood and tissues. After his Ph.D. in Biophysics (1955), Ed became a Principal Investigator for the U.S. Army Biological Laboratories at Fort Detrick, Maryland, where he continued to investigate the dielectric properties of biological tissues.

In 1961, Ed came to the University of Rochester as an Associate Professor of Electrical Engineering. He became director of one of the first National Institutes of Health-sponsored graduate training programs in the emerging field of biomedical engineering, and he continued as Director of Biomedical Engineering through 1983. Ed's early dedication and leadership set the path for what today is a vibrant Department of Biomedical Engineering at Rochester. In 1986, Ed founded and became the first Director of the Rochester Center for Biomedical Ultrasound, consisting of nearly 100 scientists, engineers, and clinicians involved in ultrasound research. Throughout his career, Ed has made outstanding and wide-ranging contributions to the field of biomedical ultrasound, some of which are highlighted below.

Continuing his pioneering work on ultrasound absorption, Ed was among the first to demonstrate the use of ultrasound for noninvasive thermal tissue destruction. Working with Charles Linke, M.D., Ed used high intensity ultrasound fields to produce thermal lesions in tumors without accelerated metastasis. These early contributions provided a foundation for high intensity focused ultrasound (HIFU) surgery. Working with high intensity fields, led Ed to observe puzzling limitations in field intensity. He had encountered acoustic saturation. This began his long interest in the nonlinear propagation of ultrasound in biological tissues. In collaboration with Tom Muir and David Blackstock at the University of Texas at Austin, Ed demonstrated the importance of nonlinear propagation in diagnostic ultrasound. His contributions set the stage for the later development of harmonic imaging in diagnostic ultrasound.

Ed has made many seminal contributions to our understanding of the physical mechanisms for effects of ultrasound in biological tissues. In cooperation with the Departments of Radiology and Medicine/Cardiology, Ed contributed to the early use of contrast agents in echocardiography. Although contrast in ultrasound then lay dormant for years, it is now a major component of echocardiography. Working with biological systems containing gas, Ed was among the first to demonstrate bioeffects of cavitation *in vivo*. After the introduction of clinical lithotripsy in the early 1980's, Ed's laboratory was among the first to characterize lithotripter fields and determine thresholds for bioeffects of acoustic shock waves in many tissues. Following his discovery that

lithotripter fields can produce lung hemorrhage, Ed was the first to show that mammalian lung hemorrhage can result from diagnostically relevant pulsed ultrasound. Subsequently, he investigated bioeffects of acoustic cavitation in mammalian systems with and without microbubble contrast agents. Today, as Senior Scientist and Yates Professor of Engineering Emeritus at Rochester, Ed continues to contribute to furthering the use of ultrasound in medicine.

Ed is a gifted and patient teacher and mentor. Throughout his academic career, he supervised six doctoral students and several Master's students, and inspired many undergraduates by welcoming them into his laboratory. Ed is a trusted colleague who generously shares his time and knowledge, and quietly influences the research paths of many of us in the field of biomedical ultrasound. Ed's outstanding achievements have been recognized widely. He was elected to the National Academy of Engineering for contributions to understanding the biological effects of ultrasound and extremely low frequency electric fields. He is a Fellow of both the Acoustical Society of America and the American Institute of Ultrasound in Medicine, which bestowed on him its Joseph H. Holmes Pioneer Award. A Life Fellow of the Institute of Electrical and Electronics Engineers (IEEE), Ed received its Engineering in Medicine and Biology Society's Career Achievement Award. Ed has been a prolific writer, authoring a book, many book chapters, and about 200 technical papers.

Ed is a devoted husband, father, grandfather, and great-grandfather. Pam and Ed have raised five exceptionally talented children, all successful across a diverse range of occupations. The family has grown to include seven grandchildren and two great-grandchildren Ed's devotion to documenting his family's genealogy will ensure that his grandchildren are well acquainted with no less than 3500 ancestors!

The ASA's Helmholtz-Rayleigh Interdisciplinary Silver Medal is awarded in highly deserved recognition of Edwin L. Carstensen's vast contributions to our understanding of the physics of biomedical ultrasound.

DIANE DALECKI
DAVID T. BLACKSTOCK
FREDERICK KREMKAU

GOLD MEDAL
of the
Acoustical Society of America

Katherine S. Harris

2007

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	Arthur H. Benade	1988
Floyd A. Firestone	1955	Richard K. Cook	1988
Harvey Fletcher	1957	Lothar W. Cremer	1989
Edward C. Wente	1959	Eugen J. Skudrzyk	1990
Georg von Békésy	1961	Manfred R. Schroeder	1991
R. Bruce Lindsay	1963	Ira J. Hirsh	1992
Hallowell Davis	1965	David T. Blackstock	1993
Vern O. Knudsen	1967	David M. Green	1994
Frederick V. Hunt	1969	Kenneth N. Stevens	1995
Warren P. Mason	1971	Ira Dyer	1996
Philip M. Morse	1973	K. Uno Ingard	1997
Leo L. Beranek	1975	Floyd Dunn	1998
Raymond W. B. Stephens	1977	Henning E. von Gierke	1999
Richard H. Bolt	1979	Murray Strasberg	2000
Harry F. Olson	1981	Herman Medwin	2001
Isadore Rudnick	1982	Robert E. Apfel	2002
Martin Greenspan	1983	Tony F. W. Embleton	2002
Robert T. Beyer	1984	Richard H. Lyon	2003
Laurence Batchelder	1985	Chester M. McKinney	2004
James L. Flanagan	1986	Allan D. Pierce	2005
Cyril M. Harris	1987	James E. West	2006

CITATION FOR KATHERINE SAFFORD HARRIS

. . . for pioneering research and leadership in speech production and dedicated service to the Society.

SALT LAKE CITY, UTAH • 6 JUNE 2007

Katherine Harris is a New England native who still owns a home in Massachusetts. She lived in other places in her youth, including Greenwood, Mississippi, where one of her high school classmates was James Flanagan. Kathy currently lives in Brooklyn, New York, where she and her husband George have resided for more than 50 years. She received a B.A. in Psychology from Radcliffe College in 1943 and a Ph.D. in Psychology from Harvard University in 1954, the year in which she joined the Acoustical Society of America.

Kathy has been recognized as an outstanding scientist in speech communication research since the beginning of her career. Specifically, Kathy's leadership in the area of speech production has been acknowledged from the beginning of her career. At the National Institutes of Health (NIH) she has participated on grant review panels, as well as serving as a consultant on task forces and advisory committees for the National Institute on Deafness and Other Communication Disorders. For 20 years she was the principal investigator on an NIH Clinical Research Center Grant, directing a large number of investigators and diverse research projects. Notably, Kathy has been named a Fellow by four distinguished societies, the Acoustical Society of America, the American Association for the Advancement of Science, The American Speech-Language-Hearing Association, and the New York Academy of Sciences. Given this outstanding research career, Kathy has frequently been invited to present her research and reviews of the spectrum of speech production research at major national and international meetings.

The word ''pioneer'' aptly characterizes Kathy's career. Kathy was a pioneer in the use of electromyography (EMG) to study the dynamics of motor control, specifically with regard to the muscles used in speech production. The fact that no procedures existed to process and reduce the data for interpretation at a time when computers were just being introduced in speech research did not intimidate her. She cheerfully lead a team of engineers, programmers, technicians, ENT physicians, fellow scientists and students to find ways to solve the technical problems and obtain useful data. In spite of recalcitrant electrode placements, disk crashes and working late-late nights, Kathy was always there with a cheery smile and a lilting ''hello'' to keep everyone going. The significant role she has played in this speech production research was acknowledged by her selection to lead the history lectures in Speech Communication in preparation for celebrating the 75th anniversary of the Acoustical Society of America (ASA) in 2004 (Harris, Ladefoged and Stevens, 2000).

Kathy was also a pioneer role model for women scientists, particularly in ASA. In the 1950s few women were active members in ASA or indeed visible in many of the hard sciences. Kathy, however, made major scientific contributions through her research at Haskins Laboratories. When she moved into her academic role as a professor in the mid-1960s, she brought her enthusiasm and dedication to science with her. This attracted a host of students, both female and male, to her and to her research. The women, however, especially benefited from Kathy's mentorship. Because she is a woman who is also a wife and mother, women students could see that practicing exemplary science and being female were not mutually exclusive.

Through Kathy's mentorship many of the field's most productive men and women have forged their own excellent careers in speech science. Both nationally and internationally, at universities and in laboratories, Kathy's students and research associates can be found among the leading figures in speech research. It is fair to say that there is no such thing as an ''ex-student'' of Kathy Harris. Indeed, her influence extends, through her students, to today's younger generation of speech scientists. This is in part due to the textbook Kathy has written with her former doctoral students Gloria Borden (Professor, Temple University) and Lawrence Raphael (Professor, Adelphi University), one of the most widely used textbooks in speech science for 20 years, the *Speech Science Primer*.

Kathy has had a life-long commitment of service to the Acoustical Society. She served as Chair of the Speech Communication Technical Committee from 1963-1965, and has served on the Committee on Medals and Awards, and the Committee on Special Fellowships. Kathy was a member of the Executive Council (1980-1983) and Vice President (1990-1991). Kathy's contributions to Speech Communication have been recognized by the Society in her election to Fellowship in the ASA (in 1967) and her receipt of the Silver Medal in Speech Communication (in 2005).

Kathy's most recent service to the Society was as President (2000-2001). Beginning in her year as President-Elect (1999-2000) and continuing into her year as President, the Society's Standards activities faced substantial challenges, including international concerns that the Society was no longer committed to standards work. These concerns had to be addressed as the Spring 2001 meeting in Vienna of the International Organization for Standardization Technical Committee 108 approached. Although somewhat tangential to her own technical area, Kathy understood that as president she had to reinforce the Society's commitment to standards work. Thus, Kathy attended the meeting as part of the US delegation. Kathy's attention to the Standards program illustrates her commitment to the Society: whatever the issue, she approaches it enthusiastically, learns everything she can, pays attention to the details, and works toward a solution.

The award of the Gold Medal is given in recognition of her life-long devotion to science and to the Society. She has earned an international reputation as a leader in speech science by contributing her insights to a variety of topics in basic and applied speech communication. Moreover, her bright disposition and inquiring mind has inspired many of the best speech scientists today to follow her example.

DIANE KEWLEY-PORT
LAWRENCE J. RAPHAEL
FREDERICA BELL-BERTI

Session 4aAA

Architectural Acoustics: Troubleshooting in Room Acoustics

Robin Glosemeyer Petrone, Chair
Jaffe Holden, 1453 3rd St. Promenade, Suite 335, Santa Monica, CA 90401

Chair's Introduction—8:45

Invited Papers

8:50

4aAA1. Acoustic enhancement in large indoor volumes. Steve Barbar (LARES Assoc., 30 Dunbarton Rd., Belmont, MA 02478)

Large indoor spaces present many unique challenges for controlling sound. If both speech and music must be simultaneously delivered with high fidelity in a large indoor volume, the challenges for sound delivery can become daunting. Carefully designed acoustical treatments and electro-acoustic systems can yield good speech intelligibility. These treatments, however, can be detrimental to the delivery of music. Integrating time-variant electronic acoustical enhancement allows both acoustical treatments and electro-acoustic systems to be used more effectively. This presentation will discuss the electro-acoustic systems installed at the conference center for the Church of Jesus Christ of Latter Day Saints. Impulse response measurements indicate how portions of each system and architecture contribute to the listening experience at different locations in the venue.

9:10

4aAA2. The use of directional and binaural microphones for troubleshooting room acoustic problems. Peter Mapp (Peter Mapp & Assoc., 101 London Rd, Copford, Colchester, Essex C06 1LG, UK, Petermapp@btinternet.com)

A common problem encountered when troubleshooting room acoustic issues (apart from the room exhibiting an overly long reverberation time or inadequate sound insulation) is the generation of late arriving reflections and echoes. The paper describes a number of techniques using both directional and binaural microphones as well as microphone array techniques for determining the direction, timing, and relative strength of such reflections. It is shown that in many situations, apparently highly directional microphones such as hypercardioid or rifle types, do not offer sufficient directional discrimination to be of much practical use. It is also shown that the inconsistent off axis frequency response of many such microphones can also lead to misleading analyses. Parabolic dish reflector microphones are shown to offer superior performance as do some microphone array techniques. It is shown that binaural measurements can also be a useful approach particularly with regard to assessing the subjective affect of reflections. The results of some early research into a novel extension to the Speech Transmission Index (STI) for assisting with acoustic trouble shooting will also be presented.

9:30

4aAA3. Interesting acoustics in the Salt Lake Tabernacle and the Latter Day Saints conference center. Sarah Rollins (Yantis Acoust. Design, 720 Olive Way, Ste. 1400, Seattle, WA 98101)

The Salt Lake Tabernacle and LDS Conference Center have interesting acoustical properties. Due to the elongated domed ceiling of the Tabernacle, there are audible focusing effects throughout the hall. In order to characterize the Tabernacle's unique acoustics prior to its seismic renovation, the acoustical properties of the hall were investigated through impulse response measurements and computer modeling. This paper will show how the measurements and model were used to see the focusing effects in the Tabernacle. The acoustics of the LDS Conference Center have also been extensively measured. This paper will show various methods that were used to investigate the sound focusing effects that were audible in the Conference Center.

9:50

4aAA4. Flutter echo at Pigott Auditorium. Basel Jurdy (Yantis Acoust. Design, 720 Olive Way, Seattle, WA 98101)

A flutter echo was easily apparent after completing the remodel of Pigott Auditorium at Seattle University in Seattle, Washington. Knowing that this was a possibility in the remodel design, walls were articulated in an attempt to avoid this phenomenon. The techniques used to investigate the problem, results of the investigation and the eventual remedy are presented.

4a THU. AM

10:30

4aAA5. Programming remediation to a newly complete performing arts center. Carl J. Rosenberg, Jennifer Hinckley, and Jonah Sacks (Acentech Inc., 33 Moulton St., Cambridge, MA 02138, crosenberg@acentech.com)

A large new performing arts building at a university included an 800-seat recital hall, proscenium theater, black box theater, large instrumental rehearsal room, choral rehearsal room, and practice rooms. It was designed with attention to acoustics concerns throughout, but the final results left the owner with a myriad of problems. These problems included poor hearing conditions in the large performance space, intrusion of exterior environmental noise, excessive mechanical system noise and vibration, and inexplicable lack of separation between rehearsal rooms that were separated by an STC 70+ wall. The authors parsed through the complaints and field conditions to extract a range of reasons for the problems. The paper presents the analysis techniques for these challenges and shows options that were developed for practical solutions.

10:50

4aAA6. Computer-aided long distance troubleshooting. Andrew N. Miller and Robert C. Coffeen (School of Architecture and Urban Design, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045)

A three-dimensional model of a religious worship facility was developed through the use of the EASE software package. Using the EASE Aura module, a binaural impulse response was created for three listening locations in the space. EASERA software was then used to analyze the impulse response data, and potential solutions to the problems observed were proposed. Each solution was then implemented in the model, and impulse responses were made for each scenario. Analysis after all proposed solutions were implemented in the model suggested that their real-world counterparts would prove successful. This paper raises the discussion of the effectiveness of long distance troubleshooting using modeling and analysis software.

11:10

4aAA7. Troubleshooting the Vivian Beaumont Theatre. J. Christopher Jaffe and Russell Cooper (Jaffe Holden Acoust., 114A Washington St., Norwalk, CT 06854)

Troubleshooting room acoustics can vary from the sublime to the ridiculous. This paper discusses some simple solutions to problems related to a lack of knowledge of basic acoustic characteristics. However, the main example in the paper relates to modifications made to the Vivian Beaumont Theatre at Lincoln Center over a period of many years. The Beaumont is a 1500-seat thrust theatre designed in the 1960s to house the Lincoln Center Repertory Company. Since opening night the theatre was plagued with poor speech intelligibility based on the facts that the seating capacity is too large for the thrust format and that the designers placed an immense proscenium stage house behind the thrust. The acoustic solutions implemented were based on discussions with two different managing directors, Joseph Pape and Bernard Gersten, and their different visions on how to mount productions in the space.

Contributed Papers

11:30

4aAA8. Lessons learned from room acoustics design and investigation (especially from using EASE room acoustics software) in our consulting practice. Joseph Bridger, Noral Stewart, and Aaron Farbo (Stewart Acoust. Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607, joe@sacnc.com)

Experiences troubleshooting during the design of new buildings and investigation of existing acoustical problems in existing buildings, including churches, dining spaces auditoria, public meeting spaces, offices, and rehearsal spaces. Lessons learned and experiences with tools, such as EASE room acoustics software. Theoretical concepts to consider in designing box shaped rooms, cafeterias, rehearsal spaces, and churches. Tools and methods to consider in acoustical investigations.

11:45

4aAA9. Case studies of troubleshooting room acoustics problems using critical listening, field measurements, and computer models. Gary W. Siebein (Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611-5702), Robert M. Lilkendey, Hyun Paek, and Edwin S. Skorski (Siebein Assoc., Inc., Gainesville, FL 32607)

Combinations of critical listening, field measurements of impulse responses and their associated metrics, and computer model studies were used in four projects to troubleshoot acoustical problems perceived by users and to study potential design solutions. Case study 1 is a 56-person firing range at a law enforcement training center nicknamed the Thunder Dome because of the excessive build-up of reflected and focused sound in the room. Case study 2 involved the diagnosis and correction of excessive loudness in a band practice room where instructors were threatening legal actions for hearing compensation claims against the school board. Case study 3 is a large multi-purpose performing arts center that suffered from many architectural acoustics and sound system design defects. Case study 4 involved the diagnosis and correction of sound focusing in an exclusive restaurant at a major resort that resulted from very subtle wall and ceiling shapes. The case studies illustrate the range of issues that can be studied with computer models and field measurements, as well as understanding the limitations of the techniques at the present time.

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration: Coded Excitation

Jonathan Mamou, Cochair

Riverside Research Inst., F. L. Lizzi Center for Biomedical Engineering, 156 William St., New York, NY 10023

Michael L. Oelze, Cochair

Univ. of Illinois, Electrical and Computer Engineering, Beckman Inst., 405 N. Mathews, Urbana, IL 61801

Invited Papers

8:30

4aBB1. An acoustic spread spectrum technique for the study of outdoor noise propagation. David C. Waddington and Jamie A. Angus (Acoust. Res. Ctr., Univ. of Salford, Salford, M5 4WT, UK, d.c.waddington@salford.ac.uk)

An investigation of the properties of a directly carrier-modulated code sequence modulation signal is presented for the study of outdoor propagation. The audible frequency carrier is biphase-modulated using a pseudorandom code sequence specially designed for simultaneous fine-time resolution and large-range ambiguity. An experiment is described in which the correlation properties of the spread spectrum signal are demonstrated, and the amplitude of the transfer functions between the receivers, together with accurate times-of-flight, are calculated from the cross-correlation of the measured acoustic spread spectrum signals. The results illustrate that an acoustic spread spectrum technique can provide significantly improved ways of measuring sound propagation outdoors. In particular, the fine-time resolution for time-of-flight suggests an application in acoustic tomography, for example in the investigation of nocturnal boundary layers, horizontal gradients, and of turbulence structures in the atmospheric surface layer. [Work supported by EPSRC UK.]

8:50

4aBB2. Complex-valued coded excitations for acoustical channel measurements. Ning Xiang and Shu Li (Grad. Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

Coded signals have been increasingly applied in acoustical/bio-ultrasonic measurements in order to increase the measurement quality or to cope with demanding requirements [Xiang and Schroeder (2003)]. Increasing interest in coded signals as acoustical excitations has prompted development of efficient correlation algorithms [Daigle and Xiang, J. Acoust. Soc. Am. (2006)]. These algorithms allow efficient processing of complex-valued sequences as needed in critical acoustical measurements, such as using simultaneous sources. Exploiting excellent correlation properties of some coded signals, this work applies complex-valued quadratic residue (QR) sequences and Zech-logarithm sequences in both base-band acoustical channel measurements using sequential excitations and band-limited acoustical channel measurements using amplitude-modulation schemes. Numerical simulations and laboratory experiments carried out in this work demonstrate feasibility of the techniques. This paper discusses advantages and challenges when using coded excitations in potential acoustical applications.

9:10

4aBB3. Enhanced spatial resolution of dual-mode ultrasound arrays using coded excitation. Yayun Wan and Emad Ebbini (Dept. of Elec. and Comput. Eng., Univ. of Minnesota, Minneapolis, MN 55455)

A prototype 64-element 1-MHz concave (ROC = 100 mm) DMUA has been shown to produce conventional synthetic aperture (SA) images in a field of view extending by 6 cm axially and 4 cm laterally (around its geometric center). With an $f_\# = 0.8$ and a fractional bandwidth $\approx 40\%$, the lateral and axial resolutions of this prototype were 1.1 mm and 2.6 mm, respectively. We have investigated the use of coded excitation and pseudoinverse filtering for enhancing the axial resolution of the DMUA. SA images were formed from a 12 μs linear chirp transmit pulse covering the frequency band of 0.5–1.5 MHz. The beamformed data was compressed using matched and regularized pseudoinverse filtering with a colored noise model. Image data were collected from a 3 D quality assurance phantom with an egg-shaped contrast target using the DMUA. The contrast ratio (CR) was 3.35, 7.8, and 6.6 for conventional pulse-echo, matched filter compression and inverse filter compression, respectively. In addition, axial and lateral speckle cell size parameters were calculated for the three methods. The axial resolution was 2.6, 2.5, and 1.65 for conventional pulse-echo, matched filter and pseudoinverse filter, respectively. Thus one can see that approximately 33% improvement in resolution was achieved without significant loss in CR or change in lateral resolution.

9:30

4aBB4. High resolution color flow—application of coded excitation in color flow imaging. Xiaohui Hao (18548 NE 57th St., Redmond, WA 98052, xiaohui.hao@ultrasonix.com)

Color flow imaging has suffered from the trade off between resolution and sensitivity ever since it was introduced. The requirement of blood flow sensitivity results in application of narrow bandwidth, long time duration pulses. The problem associated with this type of pulse is the severely degraded spatial resolution, which causes wall overwriting, blobby flow aesthetics, loss of detail flow

3153 J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America **3153**

hemodynamics, etc. Coded Excitation has been applied in B mode imaging to solve the similar tradeoff between SNR and resolution, but has not been so successful in color flow imaging. This is due to the fact that current color flow pulses are already at the regulatory limit and the focus of delivering more power for more sensitivity with coded excitation in color flow imaging. High resolution color flow imaging is an effort to introduce Chirp Coded-Excitation to achieve B-mode-like spatial resolution in color flow without sacrificing sensitivity. Wide bandwidth pulse with time bandwidth product greater than 1 is employed to bring in sensitivity/penetration and resolution. As a result, better spatial definitions, less wall overwriting, and detailed flow hemodynamics are achieved with improved flow sensitivity. Corresponding practical application issues are discussed in this paper.

9:50–10:10   Break

10:10

4aBB5. Coded excitation and subband processing for blood velocity estmation in medical ultrasound.   Fredrik Gran, Jorgen Arendt Jensen (Tech. Univ. of Denmark, Ctr. for Fast Ultrasound Imaging, Oersted-DTU, Bldg. 349, DK-2800 kgs. Lyngby, Denmark, fg@oersted.dtu.dk), Jesper Udesen, and Michael Bachmann Nielsen (Univ. Hospital of Copenhagen, Blegdamsvej 9, DK-2100 Copenhagen, Denmark)

This paper investigates the use of broadband coded excitation and subband processing for blood velocity estimation in medical ultrasound. In conventional blood velocity estimation a long (narrow-band) pulse is emitted and the blood velocity is estimated using an auto-correlation based approach. However, the axial resolution of the narrow-band pulse is too poor for brightness-mode (B-mode) imaging. Therefore, a separate transmission sequence is used for updating the B-mode image, which lowers the overall frame-rate of the system. By using broad-band excitation signals, the backscattered received signal can be divided into a number of narrow frequency bands. The blood velocity can be estimated in each of the bands and the velocity estimates can be averaged to form an improved estimate. Furthermore, since the excitation signal is broadband, no secondary B-mode sequence is required, and the frame rate can be increased. To increase the SNR for the broad-band excitation waveforms, coding is proposed. Three different coding methods are investigated: nonlinear frequency modulation, complementary (Golay) codes, and Barker codes. Code design is described for the three different methods as well as different ways of pulse compression for restoring axial resolution. The different methods were studied using an experimental ultrasound scanner and a circulating flow rig.

Contributed Papers

10:30

4aBB6. Pulse-compression ultrasound imaging using a 40-MHz annular array. Jonathan Mamou and Jeffrey A. Ketterling (F. L. Lizzi Ctr. For Biomed. Eng., Riverside Res. Inst., 156 William St., New York, NY)

High-frequency ultrasound allows fine-resolution imaging at the expense of limited depth-of-field and acoustic penetration depth. Coded-excitation imaging allows a significant increase in the signal-to-noise ratio (SNR) and acoustic penetration depth. A 40-MHz, five-element annular array with a focal length of 12 mm and a total aperture of 6 mm was made using a 9-μm thick PVDF-TrFe membrane. The transducer was excited with an optimized, custom design, 4-μs, linear chirp spanning 15 to 65 MHz that was deduced from *ex vivo* experiments and computer simulations. Images of a 12-μm wire were generated to quantify lateral and axial resolutions. All 25 transmit/receive signal combinations were digitized and post-processed for compression and synthetic focusing. Compression consisted of linearly filtering the signals with the time-reversed excitation chirp modulated by a Dolph-Chebyshev window. Results showed that resolutions were not significantly degraded when compared to a conventional monocycle excitation and that SNR was improved by more than 14 dB. Images of mouse embryos showed marked improvements in image-quality. [Work supported in part by NIH Grant R21 EB006509.]

10:45

4aBB7. Improved scatterer size estimation from ultrasound backscatter using coded excitation and pulse compression. Michael L. Oelze and Steven Kanzler (Dept. of Elec. and Computer Eng., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Scatterer size estimation from ultrasound backscatter has been used to improve the detection and classification of cancer. Estimates of scatterer size from tissues come from ultrasound backscattered signals that often have low signal-to-noise ratio (SNR). Coded excitation was explored as a means to increase the SNR of signals used in scatterer size estimation. The normalized backscatter coefficient was measured from tissue mimicking phantoms containing glass beads. The diameters of the glass beads were estimated by fitting the normalized backscatter coefficient to models for scattering from spheres [J. Acoust. Soc. Am. **23**, 405–418 (1951)]. Two

weakly-focused, single-element transducers (f0 = 5 MHz, −6 dB BW = 4.2 MHz, f/3, and f0 = 10 MHz, −6 dB BW = 9.1 MHz, f/4) were used in the measurements. For coded excitation, a linear FM chirp with a time bandwidth product of 40 was used. Three phantoms with average glass bead diameters of 30, 45, and 82 μm were scanned. Preliminary results indicated that estimate bias was reduced versus penetration depth for the coded excitation compared to conventional pulsing. In addition, estimate variance using the coded excitation technique was reduced because the SNR was increased. [Work supported by start-up funds from the University of Illinois at Urbana-Champaign.]

11:00

4aBB8. An ultrasonic imaging speckle suppression technique by means of frequency compounding and coded excitation. Jose R. Sanchez and Michael L. Oelze (Dept. of Elec. and Computer Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801)

A method for improving the contrast resolution of an ultrasonic B-mode image is proposed by combining the speckle reduction technique of frequency compounding and the pulse compression technique of resolution enhancement compression (REC). Frequency compounding is known to suppress speckle in B-mode images to provide improvements in image quality and contrast resolution, but at the expense of deteriorating the axial resolution. The REC technique is used to enhance the bandwidth of the ultrasonic imaging system by using convolution equivalence. This enhanced bandwidth translates into improvements in axial resolution of the ultrasonic imaging system. Therefore, these improvements in axial resolution are exploited to improve contrast resolution. Improvements are quantified by using contrast-to-noise ratio (CNR) metric in simulations and experiments. In simulations, a phantom with a hyperechoic target was imaged by utilizing a single element transducer with a center frequency of 2.25 MHz. In experimental measurements, a single-element transducer (f/3) with a center frequency of 2.25 MHz was used to image ±3- and ±6-dB cysts in an ATS 539 multipurpose tissue-mimicking phantom. Preliminary simulation and experimental results indicated that the CNR could be improved by up to 25% and 60%, respectively, without sacrificing axial resolution compared to conventional pulsing methods.

11:15

4aBB9. Spread spectrum sequence design for the study of outdoor noise propagation. Jamie A. Angus and David C. Waddington (Acoust. Res. Ctr., Univ. of Salford, Salford, M5 4WT, UK)

The design of carrier-modulated code sequence signals that can be used for the study of outdoor propagation, are presented. In particular, the construction of these signals such that they are robust to the time-variance of the outdoor acoustic path is discussed. The design of these signals so that they can also occupy known bandwidths is also described. The result is an audible frequency signal that is biphase-modulated by two pseudo-random code sequences in a way that is specially designed for simultaneous fine time resolution and large range ambiguity. A particular advantage of this signal is that an estimate of the time-varying Doppler can be derived by recovering the suppressed carrier from the received signal, and various means of extracting the Doppler are explained. Extensions of these signals to allow multiple simultaneous measurements are also described. [Work supported by EPSRC UK]

THURSDAY MORNING, 7 JUNE 2007 · SALON A, 9:00 TO 10:35 A.M.

Session 4aEA

Engineering Acoustics: Acoustical Measurements and Methods I

Harold C. Robinson, Chair
Naval Undersea Warfare Center, Code 2131, 1176 Howell St., Newport, RI 02841-1708

Chair's Introduction—9:00

Contributed Papers

9:05

4aEA1. Remote acoustical reconstruction of cave and pipe geometries. Todd Schultz, David Bowen, Gladys Unger, and Richard H. Lyon (Acentech Inc., 33 Moutlon St., Cambridge, MA 02138)

Military and commercial interests desire to determine the geometry of caves, tunnels, and piping systems remotely. Advance knowledge of the geometry would allow military personnel to focus on the mission objectives rather than on cave exploration and mapping. Commercial applications of remote geometry mapping include determining locations of collapsed mines and tunnels for search and rescue missions, and determining locations of breaches or constrictions in pipelines. This research effort begins to extend acoustic pulse reflectometry to these applications and only requires simple measurements made at the waveguide entrance. This implementation first measures the acoustic reflection coefficient using a modified form of the two-microphone method and computes the waveguide impulse response. Then the system uses an inverse-scattering algorithm to solve for the cross-sectional area as a function of distance. Results are shown for a simple area expansion attached to a laboratory impedance tube with a diameter of 35 mm, and for a pedestrian tunnel with a nominal square cross-section of 5.9 m^2. The impedance tube estimates match the known cross-sectional areas to within 10%. The results from the pedestrian tunnel show greater uncertainty due to frequency range limitations on the measurements, but can still capture important features of the geometry.

9:20

4aEA2. Acoustic and elastic radiative transfer theory in seismology. Jens Przybilla, Michael Korn (Inst. of Geophys. and Geology, Leipzig Univ., 04103 Leipzig, Talstr. 35, Germany, jprzybill@web.de), and Ulrich Wegler (Seismologisches Zentralobservatorium Graefenberg, D-91052 Erlangen, Mozartstr. 57, Germany)

Inversion results that were obtained by applying the acoustic and elastic radiative transfer theory to seismological data will presented. Scattering of seismic wave fields within the earth is documented by the pulse broadening of direct wave trains and the appearance of coda waves following the direct waves. The pulse broadening is caused by forward-scattered waves and the coda by wide-angle scattering. Random media are used as models for the heterogeneities of the earth. The energy transfer equations are solved with Monte Carlo methods. The results obtained with the acoustic and elastic theory for the transport scattering coefficients and absorption coefficients are compared with each other. It shows that scattering is the main process for lower frequencies. For higher frequencies intrinsic absorption is dominant. The results for elastic and acoustic radiative transfer are the same if strong forward scattering is assumed. For lower values of anisotropy the results are different.

9:35

4aEA3. An investigation into the relationship between sound intensity and normal surface velocity. Mark Boyle (School of Mech. and Aerosp. Eng., Queen's Univ. of Belfast, Ashby Bldg., Stranmillis Rd., Belfast, BT9 5AG, mboyle31@qub.ac.uk), Matthew Cassidy, Richard Cooper, Richard Gault, and Jian Wang (Queen's Univ. of Belfast, Stranmillis Rd., Belfast, BT9 5AG, UK)

The analysis of sound characteristics using boundary element methods (BEM) requires accurate sound source data of normal surface velocity or pressure as boundary conditions. However, direct measurements of normal surface velocity are not easily acquired while sound intensity measurements are relatively easy to obtain. In this study the relationship between near field sound intensity and normal surface velocity is investigated. The study first considers correlated and uncorrelated sound sources in the frequency range of 63 to 1000 Hz. A matrix of sound intensity measurements on well-distributed points in the near field of the sound source is collected. A model of the relationship between sound intensity and volume velocity is developed and used to compute the normal surface velocity of the sound source from sound intensity measurements. The normal surface velocities are then used as the boundary conditions for a BEM model of the sound source that predicts the sound intensity at the original measurement points. The measured and predicted values are compared in order to validate the model. The differences of predicted and measured sound intensities that are within ± 3 dB are considered sufficiently accurate.

9:50

4aEA4. An integrated optical microphone test-bed for acoustic measurements. Caesar T. Garcia, Neal A. Hall, F. Levent Degertekin, Baris Bicen, Kamran Jeelani, and M. Shakeel Qureshi (Dept. of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr. NW, Atlanta, GA 30332, levent.degertekin@me.gatech.edu)

A diffraction-based optical detection method for microphone applications has been demonstrated previously [Hall *et al.*, J. Acoust. Soc. Am. **118**, 3000–3009 (2005)]. This method, coupled with proper integration

4a THU. AM

techniques can produce precision measurement microphones with 24 dBA noise levels and suitable bandwidths. Thus far, these characterization studies have been performed using experimental setups, which would disturb the acoustic field due to size and non-symmetric features. In these regards, previous optical microphone test beds have been inadequate experimental platforms. This has motivated the development of a more robust integrated instrumentation microphone package for future testing and characterization. In order to meet the size restrictions for such an optical microphone platform, vertical cavity surface emitting lasers are used as light sources and small photodiode arrays are used to detect intensity variations in refracted orders of the optical detection method. The overall dimensions and shape of the package are comparable to commercially available half-inch calibration microphones and impose minimal sound field disturbance. The design is adapted to allow simple replacement and remounting for multiple microphone testing including biomimetic directional microphones [Miles *et al.*, J. Acoust. Soc. Am. **98**, 3059–3070 (1995)]. [Work partially supported by NIH Grant 5R01DC005762-03 and the Catalyst Foundation.]

10:05

4aEA5. Design and implementation of a three-dimensional seven microphone acoustic intensity probe for measuring sound properties from various sources. Khalid Miah and Elmer Hixson (Elec. and Computer Eng. Dept., Univ. of Texas, Texas 78712, miah@mail.utexas.edu)

Based on existing theoretical models of multi-dimensional cross-spectral methods for calculating sound pressure and particle velocities, this report has derived and/or extended models of intensity, energy densities, and specific acoustic impedance for seven microphone configurations. Array configuration, spacing, and overall implementation procedure is also outlined in this report. Experimental data has confirmed the validity of the analytical solution for different sound sources in both anechoic and rever-

beration room environment. Sound sources of different frequency bands [broadband (100 Hz–7 KHz), low (100 Hz–1 KHz), and high (1–7 KHz)] was used separately to compare accuracy and performance of the acoustic probe. Signal processing techniques and software implementation with potential errors and limitations of the system are explained. These would provide a framework for further refinement of instrumentation aspects of the acoustic probe. Potential applications and users of the probe are also described in this report.

10:20

4aEA6. Phonometrical precision measurements during a high end classical music live recording and four standardized recording systems' signal comparison. Franco P. A. Policardi (DIENCA, Universit di Bologna, Viale Risorgimento 2, 40136 Bologna, Italy, franz.policardi@mail.ing.unibo.it)

Discographical sound recording process is very seldom accompanied by real sound pressure, intensity and level measurements. Usually available data are often referred to mathematical extrapolations calculated on sound recording audio signals. This article investigates the situation defined as "real" from a scientifical point of view, having as a referring base a Swedish precision sound level meter fresh from calibration, during the stereophonic multichannel live recording session of the Philharmonische Camerata Berlin. Comparison among reference signal and eight high quality microphone capsules signals coming from four different stereophonic standardized microphone systems devoted to discographical purpose recording (XY, ORTF, Binaurale and AB) shows big differences among the respective signals and the same recording spot measurement system position. It is noticeable how differences even higher than 6 dB can make microphone positioning particularly critical. A standardizing method is proposed.

THURSDAY MORNING, 7 JUNE 2007　　　　　　　　　CANYON ROOM B, 10:00 TO 11:35 A.M.

Session 4aED

Education in Acoustics: Tools for Teaching Acoustics

Edward J. Tucholski, Chair
U.S. Naval Academy, Physics Dept., 572C Holloway Rd., Annapolis, MD 21401

Chair's Introduction—10:00

Contributed Papers

10:05

4aED1. An investigation of Rubens flame tube resonances. Michael D. Gardner (N283 ESC Brigham Young Univ., Provo, UT 84602), Gordon Dix (JBL Professional, Northridge, CA 91329), Kent L. Gee, and Jarom Giraud (Brigham Young Univ., Provo, UT 84602)

The Rubens flame tube is a 100+ year-old teaching demonstration that allows observers to visualize acoustic standing wave behavior [H. Rubens and O. Krigar-Menzel, Ann. Phys. (Leipzig) **17**, 149 (1905)]. Flammable gas inside the tube flows through holes drilled along the top and flames are then lit above. The tube is closed at one end and is driven with a loudspeaker at the other end. When the tube is driven at one of its resonance frequencies, the flames form a visual standing wave pattern as they vary in height according to the pressure amplitude in the tube. Although the basic performance of the tube has been explained [G. Ficken and C. Stephenson, Phys. Teach., **17**, 306–310 (1979)], a previously unreported characteristic of some flame tubes has been observed and studied. This phenomenon involves an upward shift of the natural frequencies of the lower modes from what would ordinarily be expected in a closed-closed tube. Results

from a numerical model suggest that the shift is primarily due to a Helmholtz resonance effect created by the drilled holes. The numerical results are compared to experimental findings and discussed.

10:20

4aED2. Collocated proportional-integral-derivative (PID) control of acoustic musical instruments. Edgar Berdahl and Julius Smith III (Ctr. for Comput. Res. in Music and Acoust. (CCRMA), The Knoll, 660 Lomita, Stanford, CA 94305, eberdahl@ccrma.stanford.edu)

The dynamic behavior of an acoustic musical instrument can be drastically modified by closing a feedback loop around a single sensor and actuator. Proportional-Integral-Derivative (PID) control provides for a simple yet effective paradigm, whose analysis is simplest for collocated systems. The analysis is further simplified when the musical instrument is modeled by a single, lightly damped mass-spring-damper system. This example is presented as a pedagogical laboratory exercise (http://ccrma.stanford.edu/realsimple/pidcontrol) within the framework of the Re-

alSimPLE project (http://ccrma.stanford.edu/realsimple). Students learn what physical models are, why they are useful, what the basic ideas behind feedback control are, and how they may be applied to a plucked string. A plucked string model is implemented with a digital waveguide in the Pure Data environment. Given the waveguide model and the theoretical results from the simplified mass-spring-damper model, students are asked to find PID coefficients that result in particular controlled instrument behaviors (increase in damping, decrease in pitch, etc.). At the close of the talk, a few more advanced topics are covered. In particular, a proof is presented explaining why damping with collocated velocity feedback is stable no matter how large the feedback loop gain is. [Work supported by the Wallenberg Global Learning Network.]

10:35

4aED3. Effectiveness of Physlet computer animations for enhancing student learning of acoustic principles in a course for non-science majors. Andrew A. Piacsek and Ian Wright (Dept. of Phys., Central Washington Univ., Ellensburg, WA 98926-7422)

A set of interactive computer animations of acoustic phenomena was created using the Java-based PHYSics appLETS (or physlets) developed at Davidson College [Christian and Belloni, 2001]. These animations allow the user to adjust certain parameters of a modeled system and view—in real time—the corresponding change in the physical behavior of the system. Because the animations are accessed via the web and can run on any computer (using Java), students can work with them in many different settings. The goal is to help students with limited math skills who enroll in PHYS103 "Physics of Musical Sound" at Central Washington University develop an intuitive understanding of basic acoustic phenomena and to assist them in applying quantitative reasoning to predict the behavior of musical systems. These newly-created acoustics and vibration physlets will be used for the first time during the Spring term, 2007, in the PHYS103 course taught by the author. Students will be given a quiz before and after working with the physlets on various topics to assess the effectiveness of this pedagogical tool.

10:50

4aED4. Using technology to enable people of all abilities to explore music. Alastair C. Disley, David M. Howard, and Andrew D. Hunt (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK)

MIDI Creator is a commercially available programable device that converts sensor inputs to MIDI data, which can then directly trigger sound or be used as controls for more complex environments. A variety of sensors are available, including pressure-sensitive floor pads and ultrasonic beam controllers. The authors present two applications of this technology, which illustrate the system's ability to be used in music education and music therapy environments. In the first instance, two rows of floor pads were used as controls for a game projected onto the floor. The default game was a co-operative ping-pong variant, in which users were encouraged to work together to achieve the highest number of volleys. Other options included musical games, where users could co-operate to produce harmonies. All ages and abilities, including wheelchair users, were able to take part on an equal basis. In the second instance, a simple potentiometer

was explored as an input device. School children rotating the potentiometer caused the MIDI Creator to play a scale and light lamps corresponding to the notes being played. The children then decided which notes lit which lamps. Other uses of MIDI Creator in the areas of music education and music therapy are summarized.

11:05

4aED5. Physics of a balloon lens. Derek C. Thomas, Kent L. Gee, and R. Steven Turley (Dept. of Phys. and Astron., Brigham Young Univ., N283 ESC, Provo, UT 84602, derekcthomas@gmail.com)

A balloon filled with a gas that has a different sound speed than that of air has been used by science teachers in the past as an acoustic lens. One purpose of the lens is to show refraction of sound waves as an analogy to geometric optics [H. Kruglak and C. C. Kruse, Am. J. Phys. **8**, 260–261 (1940)]. This paper discusses the physics involved with the balloon lens demonstration. In order to determine the validity of a balloon as a classroom demonstration of an acoustic lens and to understand the corresponding phenomena, the problem has been considered analytically, numerically, and experimentally. The results show that although a geometric analogy is a valid first order approximation, scattering theory is often required to fully understand the observed phenomena. Thus, this demonstration can be adapted to a wide range of students, from those learning basic principles of refraction to mathematically advanced students studying scattering.

11:20

4aED6. Combining physical REALity with SIMulations in pedagogical laboratory experiments. Edgar Berdahl, Nelson Lee, Ryan Cassidy, and Julius Smith III (Ctr. for Comput. Res. in Music and Acoust. (CCRMA), The Knoll, 660 Lomita, Stanford, CA, 94305, eberdahl@ccrma.stanford.edu)

The RealSimPLE project provides teachers with a modular collection of online musical acoustics laboratory experiments, including experiments on elementary wave motion, vibrating strings, pipes, sensors, digital waveguides, psychoacoustics, PID control, sound cards, and more. Hands-on laboratory sessions are complemented by pedagogical computer-based simulations or animations of the same systems. This dichotomous approach illustrates both the behavior of the real system together with corresponding idealized theoretical simulation. Moreover, the simulations can illustrate hard-to-measure details of the real system, or highlight details which easily escape the eye or ear. In this way, the traditional lab bench is enhanced rather than replaced. Detailed instructions help teachers and/or students easily construct the minimal amount of low-cost laboratory equipment needed. The RealSimPLE website (http://ccrma.stanford.edu/realsimple) also includes rollover "pop-up" definitions for many technical terms. These definitions (within HTML pop-up overlays) include links to more-detailed definitions and examples, allowing website visitors to easily dig deeper down through the tree of prerequisite terms, thereby filling in any knowledge-gaps they may have. A motivated student anywhere in the world with a basic math and physics background can obtain advanced graduate-level knowledge from the website in a self-paced, demand-driven manner. [This work was supported by the Wallenberg Global Learning Network.]

Session 4aMU

Musical Acoustics and Physical Acoustics: Flow Dynamics in Musical Instruments

Rolf Bader, Chair

Univ. of Hamburg, Inst. of Musicology, Neue Rabenstr. 13, 20354 Hamburg, Germany

Invited Papers

9:00

4aMU1. Turbulence flow modeling using combined compressable Euler and molecular dynamic simulation (MDS) methods.
Rolf Bader   (Insitute of Musicology, Univ. of Hamburg, 20354 Hamburg, Germany)

Typical flows found with musical instruments as well as turbulent behavior are modeled. Here the problem of using many elements to model the system precisely resulting in very long simulation time is avoided using molecular dynamic simulation techniques, where the flow is represented by moving particles that can be at any position in space and so act as an alternative to small grid lengths. On the other hand, the coupling between the flow and the acoustic wave caused by oscillations of that flow is modeled using the compressable Euler model. Here, the equation of continuation is used for modeling a time dependency of the pressure and is thus capable of combining the models. Both models are presented on their own and also in several possible combinations. The flow and pressure distribution behavior of possible couplings between these methods is then discussed in terms of comparison with experimental data, modeling stability, and modeling speed.

9:25

4aMU2. Turbulence: Geometrical and statistical aspects.   R. Friedrich   (Inst. of Theoretical Phys. Westfälische Wilhelms-Universität Münster, Wilhelm-Klemm-Str. 9, 48149 Münster, Germany)

Turbulent flows are characterized by the transport of energy across scales. This energy transport, which frequently is denoted as energy cascade, is related to the formation of hierarchically organized spatio-temporal structures. A statistical treatment of velocity increments of turbulent ows thereby shows that the statistics are strongly intermittent exhibiting large deviations from Gaussian statistics. In this talk we shall give an overview on recent results concerning the analysis of Eulerian as well as Lagrangian velocity statistics, which allows one to characterize, understand, and model the small scale statistics of turbulent flows by means of stochastic processes. Furthermore, we review attempts to relate the statistical properties of small scale turbulent flows to the geometry of vortical structures in the flow field.

9:50

4aMU3. Flutelike instruments: A perfect joint between acoustics and hydrodynamics.   Benoit Fabre   (Inst. Jean Le Rond d'Alembert, Univ. Pierre et Marie Curie, 11, rue de Lourmel, F-75015 Paris, France)

The oscillation in flutelike instruments results from the aeroacoustic coupling of an hydrodynamic mode of a jet with an acoustic mode of a resonator. Since musical instruments are empirically optimized oscillators, the modeling of flutes appears as a test bench for different approaches. Complementary descriptions have been developed in the literature from the point of view of acoustics or from the point of view of hydrodynamics. Unfortunately, the basic hypothesis in both fields are not easily compatible: for instance, as a first approximation, pipe acoustics can be studied neglecting the influence of mean flow, while the jet instability can be studied neglecting the influence of compressibility. In the presentation, we will develop an aero-acoustic approach to the oscillation in flutelike instruments, that focuses on the different approximations that need to be carried. We will emphasize on the limited range of control parameters that the current models may describe. The different elements will be discussed as well as the traditional custom to study the system as a lumped model.

10:15−10:30   Break

10:30

4aMU4. The acoustic consequences of time-dependent changes of the vocal tract shape.   Brad Story   (Univ. of Arizona, P.O. Box 210071, Dept. of Speech, Lang., and Hearing Sci., Tucson, AZ 85721)

During singing or speech production, the human vocal tract is a nearly continuously changing conduit through which sound propagates. As the vocal tract shape changes, its resonance frequencies also continuously vary, shaping an excitation signal into a sequence of vowels and consonants. This presentation will focus on modeling the time-dependent movements of the vocal tract shape and their acoustic consequences. Based on data collected with MRI and x-ray microbeam techniques, a kinematic model of the vocal tract area function has been developed that allows for efficient specification of time-dependent cross-sectional area changes in an acoustic waveguide. When coupled with a voice source, the result is a basic simulation of the sound production process from which pressures and airflows are generated. The components of the model will be presented and then used to demonstrate some time-dependent relations between the vocal tract shape and resulting acoustic characteristics. [Work supported by NIH R01-DC04789.]

3158   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3158**

10:55

4aMU5. Synchronization of organ pipes as a paradigm for nonlinear acoustical oscillations. Markus Abel and Steffen Bergweiler (Dept. of Phys., Universität Potsdam, Am Neuen Palais, Postfach 601553, D-14415 Potsdam, Germany)

From measurements on organ pipes, it has been known for a long time that the so-called Mitnahme Effekt can lead to a mutual influence of organ pipes by each other. The same holds for external driving of pipes by acoustical sources of well-defined frequencies. A locking of two differently tuned pipes to a single frequency, or to the frequency of the external driving, respectively, is observed. We measure the dependence on the coupling by varying the distance of the pipes. In great detail, we measure the dependence on the signal strength in the case of the external driving. As a result, we find a so-called Arnold tongue over a range of 50 dB, i.e., for extremely weak driving, one still finds synchronization to a common frequency. We apply the very general theory of synchronization from the point of view of nonlinear vibrations and apply it to the experimental data from organ-pipe measurements. Surprisingly, it is possible to capture the main features of the phenomenon by nonlinear oscillator equations, coupled to each other. A full explanation includes a good understanding of the aeroacoustics producing the sound of an organ pipe.

THURSDAY MORNING, 7 JUNE 2007   CANYON ROOM C, 8:00 TO 11:40 A.M.

Session 4aNSa

Noise: Environmental Noise in National Parks

Nancy S. Timmerman, Chair
25 Upton St., Boston, MA 02118-1609

Invited Papers

8:00

4aNSa1. Acoustic monitoring and resource management in national parks. Kurt Fristrup (Natural Sounds Program, Natl. Park Service, 1201 Oakridge Dr., Ste. 100, Fort Collins, CO 80525)

The Organic Act of 1917 and other legal mandates and authorities require the NPS to protect and restore park resources. Acoustic monitoring has been conducted in 43 park units. These studies have documented sound levels (in $\frac{1}{3}$ octave bands) and identified sources of noise. Analyses show that backcountry sites can be extraordinarily quiet, but noise events occur at all locations. It is clear that noise presents two threats to park resources: chronic degradation of listening opportunities as well as acute effects from prominent events. The NPS is developing metrics that assess the extent to which noise degrades the opportunities for visitors to experience the sounds intrinsic to the park unit. These metrics also apply to the degradation of aural awareness for wildlife, which can affect foraging success, predator avoidance, and acoustic communication.

8:25

4aNSa2. Modeling noise and visitor use in national parks. Kenneth J. Plotkin (Wyle Labs., 241 18th St. S., Ste. 701, Arlington, VA 22202), Steven R. Lawson (Virginia Polytechnic Inst. and State Univ., Blacksburg, VA 24061), and Kurt M. Fristrup (Natl. Park Service, Fort Collins, CO 80525)

The natural soundscape in National Parks is degraded by the intrusion of human-made noises. One usually thinks of the intrusive sounds as those from mechanical devices: aircraft, snowmobiles, off-highway vehicles, tour buses, generators, etc. Noise Model Simulation (NMSim) developed for the Park Service was designed to address those sources. There is, however, often a significant nonmechanical human noise source: the humans themselves. Visitors traveling in groups tend to talk to each other, and that speech can be an intrusive noise source. Newman *et al.* conducted a listening exercise in Muir Woods National Monument in 2005 to assess the nature of these human sounds. In the fall of 2006 Lawson *et al.* conducted noise and trail use monitoring in the Great Smoky Mountains National Park, which quantified visitor-generated sound levels and provided input to trail use modeling. Those data provide inputs to NMSim, so that the exposure of visitors to visitor noise and the impact of visitor noise on the soundscape can be quantified. The Smokies study will lead to a general merger of noise and trail use modeling, so there will be tools for this aspect of soundscape modeling and preservation in National Parks.

8:50

4aNSa3. Grand Canyon: Restoration of natural quiet (current status). Dickson J. Hingson (Sierra Club, Natl. Parks and Monuments Committee, 275 S. River Run Rd. #3, Flagstaff, AZ 86001, dhingson@infowest.com)

Mandated restoration efforts to reduce extensive aviation noise intrusions into the wild soundscape of Grand Canyon National Park will be reviewed, in context of underlying, protracted political controversy. This is timely, for the approaching, presidentially directed due date is 22 April 2008, under the National Parks Overflights Act of 1987, for the ''substantial restoration'' of ''natural quiet,'' pursuant to noise standards/specifications of the National Park Service, with due consideration to the FAA for ensuring safety of aircraft operations. The draft EIS and ''Preferred Alternative'' (release due spring 2007) will therefore vigorously resurface scientific review of primary and supplemental noise indicators and standards appropriate to national parks. Grand Canyon, having large wilderness/backcountry areas and renown, is a particularly iconic focus for scrutiny. The restoration will be viewed against actualities

of (1) the large number/types of aircraft operations at all altitudes above the park; (2) demanding management objectives and large size of Grand Canyon; and (3) its extremely quiet natural soundscape (which characterizes most desert and Colorado Plateau national parks.) Visitor experience values, including "wilderness character" and sound as resource are offered as ultimate concerns, often clashing with industrial tourism and its associated values of aerial perspective, speed/convenience, and economic return for tour operators.

9:15

4aNSa4. Protecting natural soundscapes with an adequate margin of safety. Richard Horonjeff (81 Liberty Sq. Rd., #20-B, Boxborough, MA 01719) and Grant Anderson (Harris Miller Miller & Hanson Inc., Burlington, MA 01803)

The natural ambient sound environment in park-like settings not only provides a masker for inappropriate sounds, but also affords different acoustic experiences to the park visitor due to its dynamic nature. At a single location, for example, the absence of wind and water movement may offer periodic opportunities to experience an environment completely devoid of audible sound. At the same location wind blowing gently through nearby conifers may offer a feeling of tranquility. In late spring the sound of water in a nearby stream may provide yet another experience. Each circumstance generates a different ambient sound environment, both for soundscape enjoyment as well as masking inappropriate sources (usually human activity). This dual role creates challenges in soundscape management and planning and makes difficult the use of a single ambient descriptor to protect a sites multiple non-concurrent soundscape resources. This paper examines soundscape impairment from a damage risk perspective by assuming independent, quasi-random arrivals of various soundscapes and human activity. Actual data from three western U.S. parks are used to establish reasonable parameter values for the analysis. The paper develops impairment risk with an adequate margin of safety in line with philosophies of the EPA Levels Document of 1974.

9:40

4aNSa5. Cleaning up 20th century acoustical trash from national parks: 10 ways to quiet our parks. Leslie D. Blomberg (Noise Pollution Clearinghouse, Box 1137, Montpelier, VT 05601-1137, npc@nonoise.org)

The history of noise in the 20th Century follows a consistent pattern, first the invention of new noise sources, then the growth in use of those sources, and finally the spread of those noises into previously quiet areas. Our National Parks have experienced that same pattern of invention, growth, and spread. Moreover, those previously quiet areas that succumbed to 20th Century noises often included our Parks and wildlands. Since most of the noises degrading National Parks are 20th Century inventions, this paper will examine the historical context of todays park noise, with an emphasis on developing policy options to quiet our National Parks.

10:05−10:20   Break

10:20

4aNSa6. A low-complexity environmental noise monitoring system for unattended operation in remote locations. Robert D. Collier (Thayer School of Eng., Dartmouth College, Hanover, NH 03755) and Kenneth Kaliski (Resource Systems Group, Inc., White River Junction, VT 05001)

There is a need to measure environmental noise in remote locations such as national parks over extended periods of time with a simple, readily deployed, robust, unattended data acquisition system. The original Dartmouth College prototype system was demonstrated in 2001 [Kaliski *et al.*, Proceedings NOISE-CON 2001, October 2001]. The improved Dartmouth system is described in a patent awarded to Dartmouth College in August 2006. The system conforms to ANSI/IEC Type I and Type II standards. It samples, processes, and stores 1-s A- and C-weighted equivalent sound level (Leq) values over a period of about 2 weeks. In addition, maximum, minimum, and statistical percentile values are stored in nonvolatile memory. The instrument is powered by a rechargeable battery backed up by a solar power converter. A communication interface provides for remote control. Separate software is provided to display statistical information and detection and classification clues based on user-selected time period events. The capabilities of the improved Dartmouth system are demonstrated in a selected measurement program.

10:45

4aNSa7. Anthropogenic contribution to the acoustic environment of Mexican spotted owls in the Gila National Forest, New Mexico. Ann E. Bowles, Samuel L. Denes (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109, abowles@hswri.com), Chris Hobbs, Kenneth L. Plotkin (Wyle Labs., Arlington, VA 22202), and Elizabeth Pruitt (GeoMarine, Inc., Hampton, VA 23666)

We monitored the acoustic environment of Mexican Spotted Owls using an array of Larson-Davis (LD) sound level meters (SLMs) from 2000 to 2005. Thirty-nine LD-820 and LD-824 SLMs were deployed in a distributed array across a 20 km×24 km area, collecting two-second time interval data as continuously as possible mid-April to July. Although the study was designed to monitor low-flying military jets, SLMs collected over 350,000 hr of usable environmental noise data. The data were summarized by time, average level, and on an energy basis. The most important identifiable sources differed depending on metric. On a time basis, the ambient was the greatest contributor, followed by biotic sources and thunder. Regional jets were unexpectedly important, accounting for 2% of the total, far more than military jets. However, military jets and thunder accounted for most of the exposure to high amplitude noise (proportion of 2-s samples with LAeq ¿ 60 dBA). In terms of cumulative sound exposure, thunder was the greatest contributor, followed by ambient noise and military jets. The sources most likely to impact owls will depend on the critical features of noise from their perspective; identifying these features is the focus of future analysis [Work supported by U.S. Air Force ACC/CEVP]

11:10

4aNSa8. Characterizing natural soundscapes and understanding human response to human-caused noise in a Hong Kong country park. Kin-che Lam, Kwai-cheong Chau, and Lawal M. Marafa (Dept. of Geography & Resource Management, Chinese Univ. of Hong Kong, Shatin, N.T., Hong Kong)

A study was undertaken in a country park of Hong Kong to characterize the natural soundscapes and determine the affective response of visitors to natural and human-caused sounds. Located on the eastern extremity of Hong Kong, the country park is easily accessible and characterized by sandy beaches, rocky shores, rugged slopes, and wooded ravines nested among silent hills and mountains. The country park is exposed to varying levels of transportation noise (motor vehicle, aircraft) and visitor-caused noise. The soundscapes of a variety of natural landscapes were characterized by sound walks and acoustical and psycho-acoustical analysis of 15-min and 24-h sound recordings. Interviews with the country park visitors revealed that they preferred soundscapes with natural and joyful attributes. A questionnaire survey was undertaken in several landscape-soundscape settings with different exposure levels of human-caused noise. The results show that while the visitors were slightly bothered by extraneous noise, the degree of annoyance was significantly determined by the natural set-ting, the background soundscape, and the type of activities engaged in. The findings provide an empirical basis to formulate management strategies of natural soundscapes and contribute to a growing body of research on the potential impacts of human-caused sound in country parks.

11:25

4aNSa9. Review, comparison, and evaluation of various environmental noise regulations in the United States. Jeff D. Szymanski and W. Brent Ferren (Black and Veatch Corp., 11401 Lamar Ave., Overland Park, KS 66211, szymanskijd@bv.com)

Environmental noise in the United States is often regulated at the city, county, and/or state level. A review of some representative urban, suburban, and rural regulations that provide objective noise limits was conducted. The sound level metrics, if specified, were reviewed for applicability to various noise sources. Ambient sound level data recorded over the last 10–15 years in urban, suburban, and rural areas of the United States were compared to the applicable limits, as well as to other criteria. Comparisons were evaluated for consistency and ambient sound levels were compared to USEPA guidelines in order to evaluate their relevance more than 30 years later. The regulations and guidelines reviewed were found to have varying degrees of applicability.

THURSDAY MORNING, 7 JUNE 2007

CANYON ROOM A, 10:20 A.M. TO 12:00 NOON

Session 4aNSb

Noise, Engineering Acoustics, and Signal Processing in Acoustics: Recent Advances in Active Noise Control I

Scott D. Sommerfeldt, Cochair

Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602-4673

Kenneth A. Cunefare, Cochair

Georgia Inst. of Technology, School of Mechanical Engineering, Love Building, Atlanta, GA 30332-0405

Invited Papers

10:20

4aNSb1. Recent developments in the use of structural actuators for decentralized active control. Stephen Elliott, Paolo Gardonio, Oliver Baumann, and Yoko Aoki (ISVR, Univ. of Southampton, Southampton, UK)

With an array of ideal, dual, collocated actuators and sensors, decentralized feedback controllers can not only be unconditionally stable, but their performance in reducing vibration and, hence, radiated sound, can approach that of a centralized LQG controller [W. P. Engles *et al.*, JASA **119**, 1487–1495 (2005)]. They also have a modular architecture that scales well, even for very large structures. This paper discusses some of the problems that arise in practice when proof-mass or piezoelectric transducers are used instead of ideal force or moment-pair actuators in such decentralized controllers. Proof-mass actuators are attractive when generating significant forces, but their natural frequency must be well below the first structural resonance frequency for stable operation. Even when this is achieved, impulsive forces due to the actuators hitting their end-stops have recently been shown to potentially be an additional source of instability. Piezoceramic actuators can give significant control in thinner structures, but as well as controlling out-of-plane vibration, they also create in-plane motion in the structure that can couple into closely located strain sensors causing instability in feedback controllers. Novel arrangements of such transducers will be described in which the effect of in-plane coupling is minimized, thus significantly improving their performance.

10:40

4aNSb2. Estimating the number of uncorrelated disturbance sources in structural systems. Suk-Min Moon (Adaptive Technologies, Inc., Blacksburg, VA 24060), Leslie Fowler (CSA Eng., Inc., Albuquerque, NM 87123), and Robert Clark (Duke Univ., Dept. of Mech. Eng. and Mater. Sci, Box 90300, Durham, NC 27708-0300)

The number of input signals used as reference inputs for feedforward control applications is limited due to cost, computational burden, input processing capability, and installation constraints. Identifying an optimal subset of reference sensors from a larger set capable of conveying the dynamics important in the performance path can result in greater performance with reduced complexity and order in the active control system. However, before determining an appropriate subset of sensors, the number of exogenous distur-

4a THU. AM

bance sources must be determined. Principal component analysis can be used to determine the number of uncorrelated disturbances acting on a structural system. Singular value decomposition of a covariance matrix of measured sensor signals is used to determine the number of independent disturbances present in the dynamic system. Limitations imposed by sample data length, path dynamics, and noise can limit the ability to resolve the number of exogenous disturbance sources. To estimate the number of secondary disturbance sources, a controller can be used to minimize the response due to the dominant signal source. It is then possible to identify the significant secondary sources. This can lead to improved estimation of the minimum number of reference sensors required for feedforward control in structural and structural acoustic systems.

11:00

4aNSb3. Singular value decomposition of plant matrix in active noise and vibration—some examples. Alain Berry, Yann Pasco, and Philippe-Aubert Gauthier (GAUS, Dept. of Mech. Eng., Univ. de Sherbrooke, Sherbrooke, QC, J1K 2R1, Canada)

In active noise and vibration control problems that involve many secondary sources and error sensors, the active control performance is largely related to the conditioning of the plant matrix (formed by the transfer functions between individual secondary sources and error sensors). The principal component transformation (or singular value decomposition) of the plant matrix is an interesting tool to extract dominant secondary paths and limit control efforts. Furthermore, it can be used in a feedforward LMS controller to prevent slow convergence due to ill-conditioning of the plant matrix, and adjust the convergence rate of individual system modes. This approach is discussed through two different applications: (1) the multi-harmonic active structural acoustic control of a helicopter main transmission noise using piezoceramic actuators; (2) the broadband, adaptive sound field synthesis using multiple reproduction sources. It is shown that the approach allows decreasing the required signal processing and limiting the magnitude of the control inputs. Furthermore, in the case of sound field reproduction, it allows a very elegant and insightful interpretation in terms of controlling independent radiation modes.

11:20

4aNSb4. Negative impedance and vibration control of beams and plates. Kenneth A. Cunefare (Woodruff School of Mech. Eng., The Georgia Inst. of Technol., Atlanta, GA 30332-0405, ken.cunefare@me.gatech.edu)

The use of piezoelectric actuators for noise and vibration control purposes is well established. Recently, however, the use of negative impedance circuits for the control of such actuators has received increasing attention. This paper considers the fundamental implications of negative impedance control when used in the context of active vibration suppression on beams and plates. It is shown that action of the negative impedance control may be related to the minimization of the reactive power delivered into the structure, thereby minimizing the standing wave field. If one takes the perspective that the vibration field on a structure is composed of direct and reverberant components, then minimization of the reactive power input leads to attenuation of the reverberant component. The reverberant response is global, such that its minimization leads to global vibration minimization.

11:40

4aNSb5. Active noise control for double glazing. Christian Carme (TechnoFirst, Parc Industriel de Napollon, 48, ave des Templiers, 13676 Aubagne Cedex, France)

Implementation of ANC on double glazing windows has been investigated through controlling the volume of the air gap between the two window panes. In past work, the major problem was the location (generally in the middle of the glass plane) and/or efficiency of the transducer generating the counter noise. A new transducer (called strip-transducer) overcomes this problem by locating the loudspeaker system in the window frame. The transducer consists of a beam like rigid membrane (less than 20 mm width and 500 mm length) attached to a longitudinal voice coil suspended in the magnetic field generated by a set of magnets located in a metallic U-shaped profile. These strip-transducers are driven by a local ANC control loop, the reference sensor being either a microphone or an accelerometer located in the emitting side and the error microphone being located on the receiving side. The control loop algorithm is based on TechnoFirsts patented cloverleaf filter, allowing a faster response than traditional X-LMS algorithms. The transducer design uses the concept of tuned membrane resonator (TMR) to reduce the drive energy needed. A prototype using this new concept is under experimentation and has shown very encouraging results.

Session 4aPA

Physical Acoustics: Nonlinear and Linear Seismology and Time Reversal in Solids

Brian E. Anderson, Chair

Los Alamos National Lab., Geophysics Group, Los Alamos, NM 87545

Chair's Introduction—8:30

Contributed Papers

8:35

4aPA1. Nonlinear oscillations of an object buried in soil. Evgenia A. Zabolotskaya, Yurii A. Ilinskii, and Mark F. Hamilton (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, hamilton@mail.utexas.edu)

A model is presented for nonlinear oscillations of an object buried in soil. The derivation is based on Lagrangian mechanics. The stress-free boundary condition on the surface of the ground is satisfied by introducing an image source pulsating in antiphase with the buried object. An equation of Rayleigh-Plesset form is obtained that describes the nonlinear volume oscillations in the elastic medium. In the linear approximation, the model yields an analytic expression for the resonance frequency. When elasticity of the soil is ignored, or when its elastic properties are presumed independent of depth, the resonance frequency decreases as burial depth is increased, tending toward a limiting value at saturation. When the shear modulus of the soil is permitted to increase with depth, the resonance frequency no longer saturates but instead increases with depth after passing through a minimum value. Perturbation theory is used to estimate weakly nonlinear effects. Attention is focused on second-harmonic and difference-frequency generation. [Work supported by ONR.]

8:50

4aPA2. Seismo-acoustic antivehicular landmine detection using phase data. Bill SchneckIII and Brad Libbey (Army RDECOM CERDEC NVESD, 10221 Burbeck Rd., Ft. Belvoir, VA 22060)

Landmine detection is tactically important for main supply route clearance in counter-insurgency operations. Seismo-acoustic detection is one of multiple technologies being pursued by the Army in an effort to address this need. The experiments to be presented were designed to address the use of phase data for landmine detection. These experiments are, in large part, a replication of the efforts by T Wang *et al.* [Proc. SPIE **5415**, 70–79 (2004)]. In the present work, phase and magnitude data are used to detect three targets: a V-S 1.6 anti-vehicular mine (easy target), an M-19 anti-tank mine (difficult target), and a resonant plate (very easy control). These were tested in four situations: free space, flush bury, 1 bury, and 2 bury. The soil surface velocity was measured using an array of four laser Doppler vibrometers, and the acoustic pressure used for excitation was recorded with a microphone. In certain instances, the relative phase between the surface vibration and acoustic pressure directly above the mine was different from soil alone.

9:05

4aPA3. Nonlinear acoustic landmine detection: Comparison of tuning curve experiments profiling on and off the target. Sean A. Genis and Murray S. Korman (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402)

Recent experiments [Proc. SPIE **6217**, 62170Y, Detection and Remediation Technologies for Mines and Minelike Targets XI; J.T. Broach, R.S. Harmon, J.H. Holloway, Jr.; Eds. (2006)] have shown that the soil vibrating over a drum-like acrylic landmine simulant is more nonlinear on the target than well off (due to soil alone). Here, a Laser Doppler Vibrometer (LDV) is configured to scan laterally across the buried simulant. The 10 cm diameter, 5 cm high rigid simulant has a flexible clamped 0.8 mm thick top plate and is buried 4 cm deep in dry, sifted masonry sand in a rigid soil container placed inside the anechoic chamber. At each location the sand is subjected to a series of increasing amplitude acoustic pressure level sweeps (across a 50–200 Hz bandwidth) that includes the resonant behavior. The soil surface particle velocity vs. frequency tuning curves, from locations "off target" show relatively small nonlinear effects, while those from locations "on target" exhibit clearly identifiable backbone (peak amplitude vs. corresponding resonant frequency) curves that exhibit softening. Profiles of the backbone slope vs. position indicate an extremely large contrast ratio between "on" and "off" locations compared to profiles of surface vibration obtained near resonance. [Work supported by ONR.]

9:20

4aPA4. Nonlinear time reversal acoustic method of land mine detection: Experiment and modeling. Alexander Sutin (Artann Labs., West Trenton/Stevens Inst. of Technol., Hoboken, NJ), Brad Libbey (U.S. Army RDECOM CERDEC NVESD, Fort Belvoir, VA 22060), and Armen Sarvazyan (Artann Labs., West Trenton, NJ 08618)

The nonlinear vibration response of land mines is an effective detection clue. This response can be measured by application of two frequency acoustic waves producing a vibration spectrum above the mine that is different from its surrounding. As an alternative to this approach, this work investigates time reversal techniques as a method of acoustically exciting nonlinear mine vibrations over a broad frequency range. The phase-inversion method used here requires two short broadband time-reversed focused signals with opposite sign, i.e. phase inverted. The focused signal and the inverted focused signal are broadcast sequentially and the responses are summed in post processing. This cancels the linear response leaving the nonlinear component. A theoretical model has been developed that treats the mine stiffness as a nonlinear spring. It predicts that the dominant nonlinear frequencies correspond to the second harmonics of the mine's primary resonance. Experiments were conducted using six loudspeakers in a box placed directly over a mine. The measured responses confirm most of the model predictions, and illustrate differences between the soil and mine nonlinearities. Work was supported by the U.S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate

9:35

4aPA5. Prediction of seismic pulses from an outdoor explosive source. Shahram Taherzadeh (Faculty of Technol., The Open Univ., Milton Keynes, MK7 6AA, UK) and Keith Attenborough (The Univ. of Hull, Hull, HU6 7RX, UK)

Near-surface layering of ground soil can influence propagation of seismic pulses originating from above-surface sources. Furthermore, in a cold climate snow and frozen ground can add to this layering effect. A numerical model called Fast Field Program for Layered Air Ground Systems (FFLAGS), developed originally for continuous sound sources above a porous elastic ground, is used to predict radial and vertical seismic signals from an above ground explosive charge recorded by a geophone. Simul-

3163   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3163**

4a THU. AM

taneous acoustic air pressure recording of the same charge is used to make a prediction of the source pulse shape. This procedure produces an effective linear source. In other words, nonlinear effects are assumed to be small at the ranges of interest here. Subsequently, suitable optimization methods were used to predict a set of best fit parameters for the near-surface ground structure. [Work supported in part by US Army through its European Research Office.]

9:50

4aPA6. Angle of incidence seismic/acoustic ratios during wet and dry ground conditions. Chris A. McNeill (Dept. of Phys. and Astron., 108 Lewis Hall, The Univ. of Mississippi, P.O. Box 1848, University, MS 38677) and James M. Sabatier (The Univ. of Mississippi, University, MS 38677)

Modeling the out-of-plane seismic-to-acoustic (S/A) transfer function shows frequency and angular dependences. The amplitude, frequency, and angle of incidence analysis shows rapid shifts in the frequency locations of the transfer function maxima as the angle of incidence of the sound source becomes greater than critical angles for the deeper soil layers. Outdoor measurements of the transfer function were conducted to verify this critical angle effect. These measurements used a suspended loudspeaker as a source and vertical and horizontal component geophones as receivers to measure the angular dependence of the in- and out-plane S/A coupling ratio during wet ground conditions. The calculated S/A ratio using sound speeds and layer depth collected during dry conditions revealed critical angle effects at the test site that are similar to those in the literature. However, the S/A ratio measurements during wet conditions did not exhibit such effects. Modeling the ground as a low speed wet soil layer over a fast bottom with water table sound speeds agrees with the measured wet ground transfer function data. The wet ground condition properties will be verified and S/A ratio measurements will be made during the dry ground conditions to verify these hypotheses.

10:05–10:20   Break

10:20

4aPA7. A long-term field soil survey by acoustic techniques: Part I. Linear acoustics. Zhiqu Lu (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, University, MS 38677)

Soils are complicated, multiphase granular materials; their behavior is determined by the discrete nature of these media, external and interparticle forces, interconnected porosity, and multiphase conditions. In a real world, the properties of field soils change ceaselessly due to seasonal and climate effects. It is generally difficult or impossible to simulate field soil conditions in the lab. To study and monitor the evolution process of soil properties, acoustic techniques are invoked for a long-term field soil survey. In this study, an acoustic probe with five acoustic transducers is inserted into the ground with minimum disturbance of soils. At different depths, sound speed, temperature, moisture, soil suction, and rain precipitation are recorded continuously. It is found that temperature has a negligible effect on the sound speed, moisture plays a minor role, and soil suction is the predominant factor governing the sound speed of the shallow field soil. Terzaghi's principle of effective stress is brought to explain the observations.

10:35

4aPA8. A long-term field soil survey by acoustic techniques: Part II. Nonlinear acoustics. Zhiqu Lu (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, University, MS 38677)

Soils are nonelastic and nonclassic nonlinear materials. They present loading-history-dependent, end-point memory, and hysteretic properties, which distinguish themselves from elastic materials such as intact crystal and fluids. Based on recent development in nonclassic nonlinear acoustic

theory, a phase shift method is used to determine the hysteretic nonlinearity parameter of a field soil. In this method, the phase shift between the transmitter and receiver caused by increasing the sound level is measured and used to curve fit the equation of state that takes into account hysteresis, and to calculate the nonlinearity parameter. In a long-term survey, the hysteretic nonlinearity parameter, sound speed, temperature, moisture, soil suction, and rain precipitation are measured and their relations are studied. It is found that the nonlinearity parameter is much more sensitive than sound speed to the climate change. Soil suction is the predominant factor that affects the nonlinearity parameter. Unlike sound speed, the hysteretic nonlinearity parameter is sensitive to the variation of temperature.

10:50

4aPA9. Complex source imaging using time-reversal (TR): experimental studies of spatial and temporal resolution limits. Brian E. Anderson, Paul A. Johnson, T. J. Ulrich, Michele Griffa (Los Alamos Natl. Lab., Geophys. Group, MS D443, Los Alamos, NM 87545, bea@lanl.gov), James A. Ten Cate, and Lianjie Huang (Los Alamos Natl. Lab., Los Alamos, NM 87545)

Large earthquakes are composed of a complex succession of slip events that are nearly indistinguishable on a seismogram. The question, how does an earthquake work? remains largely unsolved. The slip events on the fault plane(s) generally take place at different spatial locations and at different times. TR wave physics can be advantageously exploited to recreate, from measured signals, a spatially and/or temporally complex sound/seismic source. An experimental study is conducted to determine the spatial and temporal resolution limitations in imaging a complex source in solids, as part of our goal to understand earthquake source complexity. TR experiments are conducted on solid blocks of different materials, such as Berea sandstone and aluminum. Arrays of piezoelectric transducers are bonded to the samples for the creation of complex spatial-temporal sources, as well as to record signals. The experimental spatial and temporal resolution limits for complex source imaging will be presented as a function of material physical characteristics (e.g., Q, modulus), as well as source signal characteristics such as pulse width, frequency and repetition rate. [This work was supported by Institutional Support (LDRD) at Los Alamos National Laboratory.]

11:05

4aPA10. Nonlinear dynamics and time reversed acoustic imaging in damaged solids. T. J. Ulrich, Brian E. Anderson, Paul A. Johnson (Los Alamos Natl. Lab., Geophys. Group, M.S. D443, Los Alamos, NM 87545, tju@lanl.gov), Alexander M. Sutin (Stevens Inst. of Technol., Hoboken, NJ 07030), and Robert A. Guyer (Univ. of Massachusetts, Amherst, MA 01003)

Our ultimate goal is locating and imaging nonlinear scatterers in solids (e.g., cracks) without *a priori* knowledge of their existence. Toward that goal, two methods have been devised that combine the spatial and temporal focusing abilities of TRA with nonlinear elastic wave spectroscopy's (NEWS) sensitivity to mechanical damage. The first method uses TRA to create large amplitude signals and induce a nonlinear response in a highly localized region on a sample surface. Repeating the process in a step-scan approach provides the means to image the sample surface/near-surface with high resolution, distinguishing nonlinear features (cracks) from the linear background (undamaged material). The second method takes advantage of TRA to focus acoustic energy from two input frequencies onto an unknown, but nonlinear, scattering source, creating nonlinear wave modulation. The time reversal mirror (TRM) array records the primary and scattering source waveforms. Filtering the recorded signals for a modulation sideband, time reversing, and rebroadcasting through the TRM focuses the filtered signal onto the source of the modulation—the crack. Experimental results demonstrating each technique will be presented. [This work was supported by Institutional Support (LDRD) at Los Alamos National Laboratory.]

Session 4aPP

Psychological and Physiological Acoustics: Topics and Methods in Environmental Sound Research

Valeriy Shafiro, Cochair

Rush Univ. Medical Center, Communication Disorders and Sciences, 1653 W. Congress Parkway, Chicago, IL 60612

Brian M. Gygi, Cochair

Acoustic Research Inst., 150 Muir Rd., 151-I, Martinez, CA 94553

Chair's Introduction—8:30

Invited Papers

8:35

4aPP1. Environmental sound research as it stands today. Brian Gygi (East Bay Inst. for Res. and Education, 150 Muir Rd., Martinez, CA 94553) and Valeriy Shafiro (Rush Univ. Medical Ctr., Chicago, IL)

Environmental sound research is still in its beginning stages, although in recent years a body of research has started to accumulate, both on the perception of environmental sounds themselves and on their practical applications in other areas of auditory research and cognitive science. However, several important issues remain unresolved, such as methodologies for environmental sound research, how to coordinate research efforts in disparate areas of environmental sound study, and how best to integrate environmental sound research into the larger body of hearing research. These issues will be discussed along with some of the current and future practical applications of environmental sounds, such as in assistive listening devices and automatic sound source recognition.

8:55

4aPP2. Multiple means of conveying information through sound: Comparisons of environmental sounds and spoken language processing using converging methodologies. Frederic Dick, Robert Leech (Birkbeck College, Univ. of London, Malet St., London, WC1E 7JX, UK), Alycia Cummings, Arielle Borovsky (UC-San Diego, La Jolla, CA), and Ayse Pinar Saygin (UCL, Queen Square, London, UK)

Environmental sounds are increasingly viewed as an attractive nonlinguistic analog for studying meaningful speech in that they can convey referential—or at least associative—information about objects, scenes, and events that unfold over time. However, environmental sounds also differ significantly from speech along other perceptual and informational parameters. These cross-domain similarities and differences have proved useful in uncovering the perceptual and cognitive divisions of labor in the developing and mature brain. Our group has directly compared environmental sound and spoken language understanding in a series of behavioral and neuroimaging studies with infants, typically and atypically developing children, healthy adults, and aphasic patients. In general, our results suggest that environmental sounds and language share many of the same processing and neural resources over the lifespan. [This research is supported by the National Institutes of Health and the Medical Research Council.]

9:15

4aPP3. Effects of training on the perception of environmental sounds with limited spectral resolution. Valeriy Shafiro (Dept. Commun. Disord. & Sci., Rush Univ. Medical Ctr., Chicago, IL 60612, valeriy_shafiro@rush.edu)

Environmental sounds play an important role in maintaining listeners' awareness of the sound sources in the environment. They can warn of potential dangers (e.g., alarms, collisions) or provide aesthetic satisfaction (e.g., babbling brook, bird song). Although research indicates that decreased spectral resolution has a negative effect on environmental sound identification, little is known about perceptual adaptation to spectrally sparse environmental sounds. This project investigated changes in the ability to identify spectrally smeared environmental sounds in response to training. In a pretest-posttest design, seven naive normal-hearing listeners, identified the spectrally smeared environmental sounds obtained with a four-channel noise-based vocoder. Five training sessions were administered between the pretest and the posttest. Overall mean identification performance improved from 33% on the pretest to 63% on the posttest, with the largest improvement of 86% obtained for the sound tokens used during training. Increases in identification accuracy were also observed for alternative tokens of the training sounds (36% difference) and sound sources not included in the training set (31% difference). Among those, the greatest improvement was found among sounds with distinct envelope patterns. The implications of these findings for perceptual learning and rehabilitation programs for cochlear implant patients will be discussed.

4a THU. AM

9:35

4aPP4. Perceptual and acoustic correlates of perceptual structures of environmental sounds. Terri Bonebright (Dept. of Psych., 7 E. Larabee St., DePauw Univ., Greencastle, IN 46135, tbone@depauw.edu) and Michael Nees (Georgia Inst. of Technol., Atlanta, GA 30332)

The present study was designed to provide insight into the perceptual structure of everyday sounds and how that structure relates to the external acoustic signal. Two large stimulus sets of 74 sounds each were used to gather sorting data, attribute ratings, and basic acoustic measurements for analysis using multidimensional scaling solutions (MDS). Correlations between and among the acoustic measurements and attribute ratings were as expected. The resulting MDS solutions with regressed vectors for attribute ratings and acoustic measurements revealed a well-defined three-dimensional perceptual structure for both stimulus sets. Similarities and differences in the scaling solutions will be discussed. Information about perceptual structure from such studies can be used by researchers to increase basic knowledge about how people perceive the relationships among everyday sounds as well as by designers of virtual reality environments to assist in developing algorithms for realistic synthesized sounds.

9:55−10:10   Break

10:10

4aPP5. Influence of auditory information on auditory-visual speed discrimination. Laurie M. Heller (Dept. of Cognit. and Linguistic Sci., Brown Univ., 190 Thayer St., Providence, RI 02912, Laurie_Heller@brown.edu)

Auditory information can exert a strong influence on multimodal perception. In the experiments described here, a video of a ball rolling in depth was paired with the sound of a ball rolling. The speed of the balls' roll in the video was independently varied from the speed of the balls' roll in the soundtrack. Total duration was held constant so that the auditory cues to speed were ongoing spectro-temporal features of the rolling sounds. In a 2IFC comparison, observers indicated which ball seemed to roll faster. In half of the trials the near-threshold visual and auditory speed information was consistent, e.g., both the video and sound sped up between the first and second interval. Performance improved as much as would be predicted by a linear model. On the other half of the trials in which the auditory and visual information were inconsistent, performance declined accordingly. [Work supported by NSF.]

10:30

4aPP6. Statistical modeling of natural auditory source event perception. Richard Pastore, Jesse Flint, and Jeremy Gaston (Dept. of Psych., Binghamton Univ., Binghamton, NY)

Challenges to understanding auditory source perception can be understood by contrast with classic psychophysics. Psychophysics studies the relationship between properties of sound and perception; by mapping changes in a single perceptual dimension as a function of controlled, systematic variation in a single acoustic property. This severely limits the degrees of freedom for both the listeners strategy and the investigators interpretation of the complex, often nonlinear, sound-perception relationships. Source perception investigation not only adds the original event as an analysis stage, but also multiple levels of complexity to every component in the analysis stages. Natural source events are complex and, although acoustic properties must reflect source event structure, variation in many acoustic dimensions of the produced sounds is determined both separately and interactively by attributes that vary with all source properties, including the one(s) investigated. Furthermore, physical acoustic dimensions seldom map directly to perceptual dimensions. Finally, listeners have many degrees of freedom in assigning importance weighting to the multitude of sound attributes, differ in their weightings, and are seldom optimal. The current paper discusses statistical modeling approaches for developing a systematic understanding of the relationships among the source-sound-perception components to event perception. [Work supported by NSF grant BCS-0213666.]

10:50

4aPP7. The object of perception in impacted sound sources: The role of perceptual accessibility and information accuracy. Bruno L. Giordano, Stephen McAdams (CIRMMT, Schulich School of Music, McGill Univ., 555 Sherbrooke W., Montreal, QC, H3A 1E3 Canada), and Davide Rocchesso (Univ. of Verona, 37134, Verona, Italy)

Two independent principles could potentially govern integration of information in the perception of a sound source. First, higher perceptual weights should be given to acoustical parameters that specify accurately and unambiguously a sound source property. Second, lower perceptual weights should be given to acoustical parameters related to less perceptually accessible information (e.g., impaired discrimination). The relevance of these principles to source perception was investigated. Hardness perception was investigated for the two objects whose interaction generates an impact sound: a highly damped hammer and a freely vibrating sounding object. Acoustical analyses identified unambiguous and relatively accurate acoustical specifiers of the impacted sound source. A first experiment assessed the ability of a listener to learn to discriminate hammer or sounding object hardness, providing a measure of the perceptual accessibility of the two source properties. A second experiment investigated hardness estimation in trained and untrained listeners, providing quantification of the relative perceptual relevance of the informational sources. A final experiment conducted on simulated sound sources further characterized weighting profiles in untrained listeners, decorrelating otherwise covarying source properties. Results concerning both mechanical and acoustical correlates suggest a relevance of both principles.

11:10−11:40
Panel Discussion

Session 4aSC

Speech Communication: Cross-Linguistic Consonant Production and Perception (Poster Session)

Shawn L. Nissen, Cochair

Brigham Young Univ., 138 TLRB, 1190 N. 900 E, Provo, UT 84602

Bruce L. Smith, Cochair

Univ. of Utah, Dept. of Communication Disorders, 390 S. 1530 E., Salt Lake City, UT 84112-0252

Contributed Papers

All posters will be on display from 8:30 a.m. to 11:30 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 11:30 a.m.

4aSC1. Changes in acoustic characteristics of /r/ following production training using electropalatography (EPG). Anna Schmidt (School of Speech Path. & Aud., Kent State Univ., Kent, OH, 44240, aschmidt@kent.edu)

Visual feedback using electropalatography (EPG) is an effective method for changing production of speech sounds in children. Differences in tongue placement and general acceptability of production have been demonstrated with EPG training for tongue to palate contact. Little is known, however, about changes in the spectral characteristics of speech sounds following training. The current study examined pre and post EPG patterns for /r/ in prevocalic, intervocalic, and postvocalic positions in sentences with five weeks of intensive EPG training. In particular, spectral characteristics of /r/ were examined for differences pre and post treatment. Subjects were four children (aged 9–10). These children exhibited distorted /r/ despite several years of traditional speech therapy but learned to produce perceptually acceptable /r/ with EPG. Spectral characteristics for /r/ productions in the same sentences by four typically developing children were used for comparison. Differences will be discussed in terms of acoustic characteristics as well as changes in tongue position.

4aSC2. Whistly fricatives: MRI and acoustic data. Christine H. Shadle, Michael Proctor, and Khalil Iskarous (Haskins Labs., 300 George St., New Haven, CT 06511, shadle@haskins.yale.edu)

Nonlinear effects in airflow during fricative production can lead to a whistly acoustic quality, which is used contrastively in a few languages and can occur due to dental appliances, but also often occurs in fluent speech. In a magnetic resonance imaging (MRI) study of American English fricatives produced by five subjects, it was noted that some of the subjects produced whistly fricatives. Articulatory and acoustic analyses will be presented to argue that small articulatory changes can produce significantly different acoustic results within the same segment class for a single subject. Lip rounding and presence of a sublingual cavity can contribute to whistling, but a single subject can produce whistly or normal frication if either of these factors is present or absent. Possibly different whistle mechanisms (edge versus hole tone) could account for some of these seeming anomalies. Strident fricatives are therefore an example of a class of speech sounds where small articulatory differences can lead to significant acoustic differences. [Research supported by NIH Grant No. NIDCD R01 DC 006705.]

4aSC3. Fourth formant dip as a correlate of American English flaps. Mary Dungan, Karen Morian, Benjamin V. Tucker, and Natasha Warner (Univ. of Arizona, Box 210028, Tucson, AZ 85721-0028)

Unlike the lower formant frequencies, the fourth formant frequency (F4) is rarely an important acoustic correlate of speech. This study shows evidence of a depression in F4 for some tokens of the American English flap. F4 may be lowered during American English /r/ (Espy-Wilson *et al.*, 2000) and during retroflex sounds of some languages (e.g., Pima, as found by Avelino and Kim, 2002), but other cues are likely to be more prominent for such sounds. The current project uses a corpus of acoustic measurements of flapped English /t, d/ (e.g., pretty, order) and shows an inconsistent but large drop in F4 at the flap consonant. This occurs predominantly near an /r/, but the F4 drop is timed to the flap, not the /r/. In some tokens, this F4 drop occurs even if the flap is reduced to an approximant, leaving few other acoustic cues to its presence. The F4 drop likely reflects a sublingual cavity resonance caused by an interaction of the tongue positions for /r/ and for a flap. This study presents results on the distribution, frequency of occurrence, and magnitude of the F4 drop. Future research will investigate the perceptual role of the F4 drop.

4aSC4. Incomplete neutralization of /t/ and /d/ in flapping environments: An analysis of the North Midland dialect of American English. Wendy J. Herd (Dept. of Linguist., Univ. of Kansas, 1541 Lilac Ln., Blake Hall, Rm. 427, Lawrence, KS 66044-3177)

This paper presents an acoustic study of alveolar flaps in American English as produced by 20 speakers of a North Midland dialect spoken in the Kansas area. Participants read three lists of 164 monosyllabic and disyllabic words and nonwords containing word-final and medial /t/ and /d/. Vowel duration differences in monosyllabic and disyllabic tokens and consonant duration differences in monosyllabic tokens duplicated previous findings in that vowels preceding /d/ were significantly longer than those preceding /t/ and in that the average duration of /t/ was significantly longer than that of /d/. However, flap duration measured significantly longer for /d/-flaps than /t/-flaps. When analyzing flap frequency, it was also discovered that females flapped significantly more often than males and that participants were more likely to flap when they were less aware of the contrast between /t/ and /d/. While previous research reported that monomorphemic and high-frequency words are more likely to be flapped than bimorphemic and low-frequency words, neither word frequency nor morphological complexity affected flap frequency in the present study. Flap

4a THU. AM

3167   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   **3167**

frequency was based on a method of distinguishing flapped from un-flapped stops on a speaker-by-speaker basis, a more accurate method than that used in previous studies.

4aSC5. A motor differentiation model for liquid substitutions in children's speech. Bryan Gick (Dept. of Linguist., Univ. of British Columbia, E270-1866 Main Mall, Vancouver, BC V6T 1Z1, Canada)

Studies of lip-jaw coordination in children have shown a lack of motor differentiation between anatomically coupled articulators in young childrens speech [Green & al. 2000, JSLHR **43**: 239–255]. A model is described in which children contending with their developing motor systems generally strive to reduce the degrees of freedom of complex anatomical structures (e.g., the tongue). The claim is pursued that segmental substitutions (e.g., /w/ replacing /r/ or /l/) are the result of specific compensation strategies which aim to simplify the complexity of the articulatory task. The proposal that gestural simplification may dictate substitution strategies for liquid consonants has been suggested previously [Studdert-Kennedy & Goldstein 2003, Language Evolution, Oxford U. Pr. 235-254]. It is proposed here that gestural simplification may be achieved via one of two basic mechanisms: gestural omission and stiffening (and hence merger), and that these two mechanisms account for all of the commonly attested substitutions for English /r/ and /l/. Supporting data are presented from ultrasound studies of: postvocalic /r/ production of an 11-month-old female speaker of English, liquid production of a group of 3–5-year-old speakers of English, and liquid production and substitutions in the speech of adolescent speakers of English with speech and hearing disorders.

4aSC6. Acoustic cues of "retroflex" and "bunched" American English rhotic sound. Xinhui Zhou, Carol Espy-Wilson (Dept. of Elec. and Computer Eng., Univ. of Maryland, College Park, MD 20742), Mark Tiede (Haskins Labs, New Haven, CT 06511), and Suzanne Boyce (Univ. of Cincinnati, Cincinnati, OH 45202)

The North American rhotic liquid has two well-known and maximally distinct articulatory variants, the classic retroflex tongue posture and the classic bunched tongue posture. The evidence for acoustic difference between them is reexamined using magnetic resonance images of the vocal tracts from two similar-sized subjects with different tongue postures of sustained /r/. Three-dimensional finite element analysis is performed to investigate the acoustic wave propagation property inside the vocal tract, the acoustic response, and the area function extraction based on pressure isosurfaces. Sensitivity functions are studied for formant-cavity affiliation. It is revealed that these two variants have similar patterns of F1–F3 and zero frequency. However, the retroflex variant is predicted to have a larger difference between F4 and F5 than the bunched one (1400 Hz versus 700 Hz). This difference can be explained by the geometric differences between them, in particular, the shorter, narrower, and more forward palatal constriction of the retroflex variant. In addition, the predicted formant pattern is confirmed by measurement from acoustic data of sustained /r/ and /r/ in nonsense words from several additional subjects. These results may prove to be useful to researchers in speech motor control, speech pathology, speaker verification, speech recognition, and speech synthesis. [Research supported by NIH.]

4aSC7. Talker-specific contextual influences on voice-onset-time. Rachel M. Theodore, Joanne L. Miller, and David DeSteno (Dept. of Psych.—125 NI, Northeastern Univ., 360 Huntington Ave., Boston, MA, 02115-5000)

Research has shown robust contextual influences on voice-onset-time (VOT) in speech production. The current work examines talker-specificity for two such cases: speaking rate (VOT increases as syllable duration increases) and place of articulation (VOT increases as place moves from anterior to posterior position). Tokens of /pi/ (labial) and /ki/ (velar) were elicited from talkers across a range of rates. VOT and syllable duration

were measured for each token. For each talker, separate labial and velar linear functions relating VOT to syllable duration were calculated. Ongoing analyses indicate that (1) for both the labial and velar functions there is significant variability across talkers' slopes [see also Theodore *et al.*, J. Acoust. Soc. Am. **120**, 3293 (2006)], but there is no significant variability in the difference between labial and velar slopes for a given talker. Thus the effect of speaking rate is talker-specific, and stable across place of articulation. (2) For each talker, the velar intercept is located at a longer VOT than the labial intercept, with significant variability in the magnitude of displacement across talkers. Thus the effect of place is also talker specific. These findings support the view that phonetic properties of speech include talker-specific contextual influences. [Work supported by NIH/NIDCD.]

4aSC8. Rate normalization of closure duration influences on an assimilation context effect. David Gow (Massachusetts General Hospital, CPZ-S340, 175 Cambridge St., Boston, MA 02114 and Salem State College, 352 Lafayette St., Salem, MA 01970) and Claire Winchell-Manning (Salem State College, Salem, MA 01970)

Variation in speech rate influences the structure and distribution of phonetic cues in a variety of ways. For example, a large body of work demonstrates that the rate of speech and nonspeech precursors influences the interpretation of temporal cues associated with phoneme categorization. Changes in speaking rate also influence the absolute size of the temporal window over which multiple cues to the identity of a segment may be distributed. This has potentially significant implications for the perception of assimilated speech in which feature cues from adjacent segments interact to produce bidirectional context effects. These implications were explored in a series of mousetracking experiments involving a non-native gradient regressive voice assimilation process that is sensitive to the duration of a silent interval between the offset of a partially voiced word-final fricative and the onset of a following voiced or unvoiced stop. Manipulations of precursor rate interacted with assimilation context effects in the interpretation of an /ais/-/aiz/ continuum. These results are discussed in the context of competing interpretations of assimilation context effects and rate normalization. [Work supported by NIH R01DC3108.]

4aSC9. Consonant duration and vowel-to-vowel coarticulation in Japanese. Anders Lofqvist (Haskins Labs., 300 George St., New Haven, CT 06511, lofquist@haskins.yale.edu)

This study examined vowel-to-vowel coarticulation in sequences of vowel-bilabial consonant-vowel, where the duration of the oral closure for the consonant is varied for linguistic purposes. Native speakers of Japanese served as subjects. The linguistic material consisted of Japanese word pairs that only differed in the duration of the labial consonant, which was either long or short; the durational ratio of short and long consonants in Japanese is about 1:2. Recordings were made of lip and tongue movements using a magnetometer system. It was hypothesized that there would be greater vowel-to-vowel coarticulation in the context of a short consonant, since a long consonant would allow the tongue more time to move. The results do not show any strong support for this hypothesis, however. The reason is that Japanese speakers modulate the speed of the tongue movement between the two vowels according to consonant duration, making it faster during a short consonant and slower during a long consonant. [Work supported by NIH.]

4aSC10. Voiced geminates in Yamagata and Tokyo Japanese. Lynnika Butler (Univ. of Arizona, Douglass Bldg. Rm. 200E, Tucson, AZ 85721, lgbutler@email.arizona.edu)

Voiced geminate stops (/bb/, /dd/, /gg/, etc.) are cross-linguistically rare, due to aerodynamic constraints on voicing during prolonged closure of the vocal tract. Standard Tokyo Japanese prohibits voiced geminate stops in native and Chinese-borrowed vocabulary, but permits them in

some recent loanwords (primarily English borrowings), where they reflect stress on the preceding syllable of the word in the originating language. (1)big–/biggu/ (2) headphone–/heddohoN/ Yachi-ben (YB) is a local sub-dialect of the northeastern (Tohoku) dialect of Japanese, which is characterized in part by voicing of most non-initial stops. Transcriptions of YB speech recorded in the 1970s by Japans National Language Research Institute (NLRI) indicate a sizeable number of voiced geminate stops, most frequently /dd/ but also /gg/ (there are no /bb/ tokens). In this study, voiced geminate stops from the NLRI recordings were compared with voiced geminate stops in English loanwords as pronounced by Tokyo Japanese (TJ) speakers. Closure duration, as well as duration and amplitude of closure voicing, were measured for both sets of data. Overall, voicing ratios (the percent of closure duration in which voicing was present) were found to be much longer in YB voiced geminates than in TJ voiced geminates.

4aSC11. Acoustic correlates of emphatic consonants in Arabic. Allard Jongman, Wendy Herd, and Mohammad Al-Masri (Linguist. Dept., Univ. of Kansas, 1541 Lilac Ln., Lawrence, KS 66044, jongman@ku.edu)

Emphasis is a distinctive feature of Semitic languages such as Arabic that refers to a group of consonants produced with a secondary constriction in the posterior vocal tract and a primary constriction typically in the dental/alveolar region. Initial findings of a detailed acoustic investigation of emphasis in Urban Jordanian Arabic will be presented. Eight adult speakers of Urban Jordanian Arabic (four males and four females) produced a list of minimal pairs of emphatic versus plain CVC words. The list consisted of four UJA consonant pairs:/d/,/d^{backglotstop}/,/t/, /t^{backglotstop}/,/ð/, /ð^{backglotstop}/, and /s/, /s^{backglotstop}/, along with the six vowels of UJA (long and short /i/, /a/, and /u/. Each target consonant occurred in both word−initial and word−final position. Results show consistently higher F1 and F3 values and lower F2 values for vowels in emphatic consonantal environments compared to vowels in plain environments. In addition, emphatic consonants have a lower spectral mean than their plain counterparts. Results from bisyllabic words will also be presented to evaluate the direction and extent of the spread of emphasis within the word. [Research supported by NSF.]

4aSC12. Production of Parisian stops word-initial stops in the vicinity of a voiceless stop. Nassima Abdelli-Beruh (Long Island Univ., C.W. Post Campus, 720 Northern Blvd., Brookville, NY 11548)

This study examines how French speakers produced stop voicing contrasts in the sentence répète CVC tout-de-suite. Durations (sentence, vowel before and after word-initial stops, closure voicing, VOT) and number of produced stop closures (/t/, /C/) were collected. There was a strong positive correlation between speaker's sentence durations and number of closure produced: Rapid speakers produced mostly one closure, whereas slow speakers generally produced two closures. One-closure and two-closure productions were separately analyzed. In the two-closure productions, significant (longer voicing in the /t/ closure in anticipation of /bdg/ than /ptk/, longer closure voicing for /bdg/ than for /ptk/, longer vowel duration following /bdg/ than /ptk/) and nonsignificant voicing-related duration differences (e.g., closure) were found. Unlike in the two-closure productions, average closure duration was significantly longer before /ptk/ than before /bdg/, while there was no significant voicing-conditioned effect of the word-initial stop on the duration of the following vowel in the one-closure productions. As in the two-closure productions, average closure voicing duration remained significantly longer before /bdg/ than before /ptk/. Findings suggest that closure voicing remains the most salient cue to the voicing distinction irrespective of contextual and speaking rate variations.

4aSC13. A preliminary investigation of voice onset time associated with Turkish stops. Janine Sadaka (Dept. of Commun. Sci. and Disord., Long Island Univ., C.W. Post, 720 Northern Blvd., Brookville, NY 11548, janinesadaka@gmail.com) and Manwa L. Ng (The Univ. of Hong Kong, Sai Ying Pun, Hong Kong)

Voice onset time (VOT) refers to the timing between the release of a stop and the onset of the following vowel. Aspiration directly reflects the period of voicelessness following a stop release. Both VOT and aspiration are temporal measurements that are results of coordinated articulatory action. The present study will attempt to determine the relationship between VOT and aspiration, and VOT and voicelessness of Turkish stops. Comparison will also be made with those reported for English. Voicing differentiation in stops requires complex motor control. Thus, many speech disorders are associated with distorted VOT. Results will be used to develop a norm on VOT for Turkish stops. This will assist in evaluating speech disorders in Turkish-speaking individuals that involve motor control problems.

4aSC14. On the phonetics of Cypriot-Greek geminate plosives. Christiana Christodoulou (2111 Lower Mall, Univ. of British Columbia, Vancouver, BCV6T 1Z4, Canada)

This research provides evidence toward the existence of geminates in Cypriot Greek (hereinafter, CyG). In the existing literature it is reported that the sole or primary cue to gemination, across languages, is closure duration. This phonetic study verifies that this cross-linguistic phonetic cue is also present in CyG. Toward this end, statistical analyses evidence highly significant durational differences in the voice onset time (VOT) of geminates versus singletons in CyG [$F(164)=p<0.0001$]. In fact it can safely be argued that VOT is the primary cue to gemination when it comes to CyG geminate plosives. This study contradicts previous claims [Arvaniti and Tserdanelis, 562 (2000); Arvaniti, 602 (1999b)] that argue closure duration to be the primary cue to gemination for CyG geminate stops. Furthermore, durational differences amongst the three voiceless plosives for both VOT and closure duration were highly evident. To the researchers' knowledge this is the first study on CyG stops that considers durational differences with respect to the place of articulation, not only for singletons but also for geminates. Preliminary statistical analysis suggests an effect of the vowel following the target segment, a fact that could facilitate in resolving the phonological representation of utterance initial geminates.

4aSC15. The effects of morphological boundaries and speech rate on geminate closures in Turkish. Molly Babel and Anne Pycha (Univ. of California, Berkeley, Berkeley, CA 94720-2650)

In Turkish, geminate consonants can occur in a word root and in derived positions across morphemes. In this experiment we explore closure duration and VOT in geminate consonants within roots, across root-suffix boundaries, suffix-suffix boundaries, and novel compound boundaries. Five (3=male, 2=female) native speakers of Turkish were recorded producing 80 target and 20 filler words with an equal number of [tt], [kk], [ll], and [mm] geminates in the four different morphological contexts named above. Subjects were instructed to produce each word in a carrier phrase three times. The wordlist was read twice by each talker. The first time subjects were asked to read at a slow, comfortable pace. The second time through, the list subjects were instructed to speak as quickly as possible without making mistakes. Consonant closure duration will be measured for the geminates. The voice onset time associated with [tt] and [kk] closures will also be measured. The results of the investigation have implications for theoretic phonological accounts of geminates.

4a THU. AM

4aSC16. Processing of fine phonetic detail in American English flaps.
Benjamin V. Tucker (Dept. of Linguist., Univ. of Arizona, P.O. Box 210028, Tucson, AZ 85721-0028)

Listeners are very good at recognizing casual, reduced speech in their native language. In day-to-day conversation acoustic information is often impoverished or reduced. Reduction in speech can occur in the form of deletion and/or change of segments, syllables, or words [Johnson (2004)]. This study explores differences in processing between reduced/conversational speech and unreduced/careful speech. Two lexical decision studies, using reduced and unreduced pairs of the American English flap [Warner and Tucker (2007)], investigate whether processing differences exist between reduced and unreduced speech. The first study presented target words to listeners in isolation and the second study presented the same targets preceded by a reduced or unreduced frame sentence. The materials contained, among 260 items, 40 word medial /d/ phonemes and 40 word medial /g/ phonemes occurring in a flapping environment [Zue and Laferriere (1979)]; targets were produced in both reduced and unreduced form. These materials, spoken by a native American English speaker, were presented to 55 monolingual American English students from the University of Arizona. Listeners performed the lexical decision task for each item and response time and error rate was measured. Native English listeners process unreduced items more quickly than reduced items.

4aSC17. Nature versus nurture: Children's development and perception on syllable-final nasals in Taiwan Mandarin.
Renee Hung (Grad. Inst. of English, Natl. Taiwan Normal Univ., No. 162, Sec. 1, HePing E. Rd, Taipei Taiwan 106, reneejh@hotmail.com) and Janice Fon (Grad. Inst. of Linguist., Natl. Taiwan Univ., Taipei, Taiwan 106)

In Hung and Fon's study (2006), it was found that elementary school children in Taiwan also tend to merge /əN/ with /ən/ as the way adults do, and the merging rates found in reading characters and sentences are higher than those in reading Zhuyinfuhao, a local phonetic system. However, while most adults and older elementary school children merge /in/ with /iN/, younger children have much lower merging rates for /i/ no matter under what condition. The first part of this study extends its investigation to preschoolers who have not officially learned Zhuyinfuhao. A picture-naming task was conducted with 20 kindergarten children. The results show that preschoolers have very similar performances with the adults' for both vowels. The second part of this study is a perception experiment. Sixty Grade-2 and Grade-4 students participated in the task, in which they were asked to write down the characters and phonetic symbols of the monosyllabic words they heard. Preliminary results indicate that these children are apt to treat /in/ and /əN/ words as homophones of /iN/ and /ən/ words, respectively, although such occurrences for /ə/ are lower than those for /i/.

THURSDAY MORNING, 7 JUNE 2007   GRAND BALLROOM C, 8:00 TO 11:50 A.M.

Session 4aSP

Signal Processing in Acoustics, Acoustical Oceanography and Underwater Acoustics: Signal Processing and Uncertainty I

Lee Culver, Cochair
Pennsylvania State Univ., Applied Research Lab., P. O. Box 30, State College, PA 16804-0030

Ning Xiang, Cochair
Rensselaer Polytechnic Inst., Architecture, 100 8th St., Greene Bldg., Troy, NY 12180

Chair's Introduction—8:00

8:05

4aSP1. Sequential Bayesian signal processing: An overview.
James Candy (Univ. of California, Lawrence Livermore Natl. Lab., P.O. Box 808, L-156, Livermore, CA 94551)

Bayesian signal processing is concerned with the estimation of an underlying posterior distribution to extract statistics (inference) solving a variety of problems (signal enhancement, detection, parameter estimation, etc.). The technique simply employs Bayes rule to estimate the posterior using all of the data available at the time. Bayesian methods require approaches that capture all of the *a priori* information available and are typically model-based employing mathematical representations of the embedded processes. An overview of Bayesian inference from batch to sequential processors is presented. Once the evolving Bayesian paradigm is established, simulation-based methods using sampling theory and Monte Carlo realizations are discussed. Here the usual limitations prevalent in classical processing algorithms (e.g., Kalman filters) are no longer a restriction to perform Bayesian inference. Next importance sampling methods are briefly discussed and shown how they can be extended to sequential solutions implemented using Markovian models as a natural evolution. With this in mind, the idea of a particle filter, which is a discrete representation of a probability distribution, is developed and shown how it can be implemented using sequential importance sampling/resampling methods. Finally, a set of applications is briefly discussed comparing the performance of the particle filter designs with classical implementations.

8:35

4aSP2. Bayesian inversion methods in ocean geoacoustics.
Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC, Canada, sdosso@uvic.ca)

This paper describes a complete approach to the inversion of ocean acoustic data for environmental model parameters, which is also applicable to other inverse problems. Within a Bayesian formulation, the general solution is given by the posterior probability density (PPD) of the model parameters, which includes both data and prior information. Properties of the PPD, such as optimal parameter estimates, variance/covariance, inter-parameter correlations, and marginal probability distributions, are computed numeri-

cally for nonlinear inverse problems using Markov-chain Monte Carlo (MCMC) importance sampling methods. Since the data uncertainty distribution (including measurement and theory errors) is generally not known *a priori*, a simple, physically-reasonable form, such as a Gaussian or double-exponential distribution, is assumed, with statistical properties estimated from data residual analysis. In many cases, the full error covariance matrix (including off-diagonal terms) is required, and in some cases effects of nonstationary errors must be included. If biased data errors are suspected, additional unknown parameters representing the biases are included explicitly in the inversion. The validity/applicability of the above assumptions and estimates is examined *a posteriori* by applying both qualitative and quantitative statistical tests. New advances in efficient and adaptive MCMC sampling for nonlinear inversion will be presented.

8:55

4aSP3. Particle filtering for dispersion curve estimation from spectrograms of acoustic signals.   Ivan Zorych, Zoi-Heleni Michalopoulou   (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102), James H. Miller, and Gopu R. Potty (Univ. of Rhode Island, Narragansett, RI 02882)

In this work we develop a particle filtering method for dispersion curve extraction from spectrograms of acoustic signals, the goal being to obtain accurate representation of modal dispersion for geoacoustic inversion. The approach combines particle filtering with modeling of sound propagation in ocean environments to track dispersion curves of multimodal signals in noisy environments. ''Crispy'' estimates of dispersion curves are extracted that are not characterized by extensive uncertainty typically hindering geoacoustic inversion. Results are presented from both synthetic signals and real data collected in the East China Sea during the ASIAEX experiment. [Work supported by ONR.]

9:15

4aSP4. Sampling with heuristics.   Kevin H. Knuth   (Dept. of Phys., Univ. at Albany, 1400 Washington Ave., Albany, NY 12222)

Algorithms that rely on heuristics are attractive to those interested in solving acoustic problems. The main reason is that heuristics often derive from solutions to a simplified problem and, as such, they are able to accurately estimate a subset of model parameters. However, algorithms that rely on heuristics are not sufficiently robust as they ignore vast regions of the parameter space. For this reason, sampling techniques that focus on exploration increase the probability of finding a difficult solution. However, sampling algorithms can become trapped in difficult problems that possess vast plateaus in the parameter space upon which the solution space occupies an extremely small volume. In these cases, heuristics at least have a chance to succeed where sampling is likely to fail. This work considers incorporating heuristically-obtained samples into a relatively new sampling algorithm called nested sampling. Nested sampling strives to localize the solution space while simultaneously integrating the posterior probability to compute the Bayesian evidence. The challenge is that nested sampling accomplishes this by working with uniformly distributed samples, whereas heuristics necessarily result in nonuniformly distributed samples. The difficulties and benefits in combining these two search strategies are described in the context of acoustic source separation.

9:35

4aSP5. The estimated ocean detector: Detection of signals with different parameter distributions.   Jeffrey Ballard, Colin Jemmott, R. Lee Culver, and Leon Sibul   (Grad. Prog. in Acoust., The Pennsylvania State Univ., P.O. Box 30, State College, PA 16804, jab795@psu.edu)

Earlier we presented a maximum likelihood receiver for acoustic signals that have propagated through a random or uncertain ocean, the estimated ocean detector (EOD) [Ballard *et al.*, Oceans 2006, Boston, MA]. In general, the EOD incorporates statistical knowledge of signal uncertainty in the form of signal parameters probability density functions (pdfs). Note that Monte Carlo simulation and possibly other techniques can utilize deterministic and statistical knowledge of the environmental to predict signal parameter pdfs. The EOD utilizes the *a priori* signal parameter pdfs to estimate signal parameters from the data (or observations), then correlates the estimate with the data; hence, the name estimator-correlator is also used. Previously we showed that, for Gaussian signals embedded in Gaussian noise, the EOD reduces to the weighted sum of an energy detector and a correlation detector. This talk presents an EOD formulated to distinguish between signals whose parameters possess different *a priori* distributions. Performance is seen to depend upon the difference between parameter distributions for the two different signals. [Work supported by ONR Undersea Signal Processing Code 321US.]

9:55

4aSP6. Calibration of source level, expected transmitted power, and reverberation for long-range sonar in fluctuating range-dependent ocean waveguides.   Purnima Ratilal, Mark Andrews, Ameya Galinde   (Northeastern Univ., 409 Dana Res. Ctr., Boston, MA 02115), and Nicholas Makris   (MIT, Cambridge, MA 02139)

The ocean waveguide is temporally and spatially varying. An acoustic signal transmitted through the ocean or scattered from objects will get randomized. Estimating parameters of sonar, the environment, and scatterers requires a statistical approach that incorporates medium uncertainties into the signal analysis. Here we discuss source level, transmission loss (TL), and reverberation model calibration using data acquired with a long-range sonar on the New Jersey STRATAFORM. Broadband acoustic data measured by a desensitized hydrophone in a towed horizontal receiving array at numerous locations from a vertical source array are analyzed. The match-filtered data are compared to the expected TL output from a parabolic equation model that accounts for bathymetric variations. A maximum likelihood estimator is implemented to provide a global inversion of the data for source level, attenuation due to scattering, and match-filter degradation in the multi-modal ocean waveguide. An estimate is also provided of the coherence bandwidth for broadband acoustic signal transmission in this environment. A challenge in calibrating bottom reverberation models with sonar data lies in separating the scattered intensity from moving objects, such as fish groups, and distinguishing them from the statistically stationary background reverberation. An approach is presented for this purpose.

10:30

4aSP7. Application of multitaper methods to passive sonar. Kathleen Wage (George Mason Univ., 4400 Univ. Dr., MSN 1G5, Fairfax, VA 22030)

While the multitaper spectral estimation method [Thomson, Proc. IEEE **70**, 1055–1096 (1982)] is used extensively in time series analysis, it has received much less attention in the array processing literature. The multitaper approach reduces the variance of power estimates by averaging windowed Fourier transforms of the data obtained using a set of orthogonal window functions. One advantage of the multitaper method is that it is designed to work with low sample support. Since multitaper estimates are often formed using a single data snapshot, they are suitable for use in highly nonstationary environments where it is difficult to obtain reliable estimates of the noise and interference statistics. This talk focuses on the application of the multitaper method to the passive sonar array processing problem. Standard techniques exist for detecting the presence of line components using multitaper estimates. This talk describes how to extend these techniques to detect plane waves in sonar data, where the signals are typically modeled as complex Gaussian rather than deterministic. Processing of non-plane-wave signals, e.g., normal modes, will also be briefly discussed. [Work supported by an ONR Young Investigator Award.]

10:50

4aSP8. Exploiting waveguide invariant reverberation striations for improved active sonar detection and classification. Ryan Goldhahn, Jeff Rogers, Granger Hickman, and Jeffrey Krolik (Duke Univ., Durham, NC 27708-0291)

Active sonar reverberation level (RL) uncertainty is exacerbated in shallow water by multipath propagation and multiple interactions with the bottom. This affects, for example, the statistics of split-window RL estimates used for cell-averaging constant false alarm rate (CA-CFAR) normalization of A-scan returns. For larger signal bandwidths, target detection is impacted by more incoherent reverberation components causing higher background variance while, alternatively, using smaller bandwidth signals results in higher range variability of the mean due to coherent multipath modal interference. In this paper, we present an alternative to CA-CFAR normalization that accounts for the frequency-selective fading characteristics of the multipath channel. The idea is to use the waveguide invariant property to estimate the frequency-dependent reverberation level at the range cell of interest using neighboring range cells at frequencies along striations in the time-frequency distribution of the sonar return. An approximate generalized likelihood ratio test (GLRT) detector is derived and demonstrated on acoustic tank data as well as Mediterranean active sonar data collected during the SCARAB-98 experiment [C. W. Holland, J. Acoust. Soc. Am. (2006)]. Detection and classification results indicate the considerable potential of a waveguide invariant-based approach. [Work supported by ONR321US.]

11:10

4aSP9. Sonar target tracking incorporating uncertain estimates of signal to noise ratio. Warren L. J. Fox, James W. Pitton, and Eric A. Rust (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, warren@apl.washington.edu)

In the standard approach to sequential estimation for target tracking, tracks are formed on contacts that exceed some predefined detection threshold, set to achieve desired P_D and P_{FA} based on an expected SNR and *a priori* data probability PDFs. If the expected SNR is erroneous, degradation will be observed as either excessive false contacts or missed detections for weak targets. A method known as the Probabilistic Data Association Filter with Amplitude Information (PDAFAI) estimates SNR from the target and interference statistics, and uses the amplitude of the contacts together with the estimated SNR in the track association stage of the tracker. This method requires the PDF for the observed signal. In active sonar, this PDF is dependent on the expected SNR, which is a spatially and temporally varying quantity. Previous work has extended the PDAFAI method by adapting the tracker to the acoustic propagation environment via the measured contact amplitude together with the expected SNR derived from an acoustic propagation model, and by incorporating target strength estimates for contacts in developing tracks. Here, the method is extended further to include probabilistic descriptions of SNR, derived both from uncertainties in estimated environmental (e.g., bottom geoacoustic) and target properties.

11:30

4aSP10. Active target tracking using the bistatic invariance principle. Lisa Zurk and Chensong He (Portland State Univ., P.O. Box 751, Portland, OR 97207, zurkl@cecs.pdx.edu)

The acoustic pressure received from a moving source in a shallow water channel is highly variable and depends critically on characteristics of the channel such as bathymetry, sound speed, and bottom properties. Past work has shown that knowledge of the channel properties can be embedded in physics-based signal processing algorithms for increased system performance. However, this requires accurate knowledge of the channel characteristics, which are often poorly known. An alternate approach is to identify aspects of the field structure that are invariant to small perturbations in the propagation environment. The waveguide invariant is a scalar parameter that has been used extensively in passive sonar to interpret features such as intensity maxima in lofargrams. Recent work has used experimental data to demonstrate that a similar parameter exists for active geometries, although the invariant relationship is more complicated because of the multiple propagation paths. In this presentation active invariant striations are exploited in an active tracker formulation. The tracker utilizes a state space representation that includes the time-dependent frequency content of the pulse as one of the sonar observables. In addition to target kinetics, the time-frequency structure provides track constraints that improve discrimination of target versus bottom reverberation. [Work sponsored by ONR-321US.]

Session 4pAAa

Architectural Acoustics: Workshop for American Institute of Architects Continuing Education Units Presentation

Bennett M. Brooks, Chair

Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, CT 06066

Invited Papers

1:30

4pAAa1. Introduction to workshop goals. Anthony K. Hoover (Cavanaugh Tocci Assoc., 327 F Boston Post Rd., Sudbury, MA 01776) and Bennett M. Brooks (Brooks Acoust. Corp., Vernon, CT 06066)

Many states require that architects obtain continuing education unit (CEU) credits annually, in order to maintain their registration or licensure. The American Institute of Architects (AIA) Continuing Education System (CES) offers continuing education courses, which may be given by third party providers. The Technical Committee on Architectural Acoustics (TCAA) is an AIA/CES Registered Provider. The goal of this workshop is to prepare members of the TCAA so that they may be authorized to present a short course that can earn attendees CEU credit. In order for TCAA members to qualify to meet the AIA requirements, they must attend this workshop, which is given in two parts. The first part will be devoted to the short course presentation material, which is in a standardized format. The second part of the workshop will focus on the AIA CEU Program registration and reporting requirements. Of course, anyone is free to register with the AIA to provide their own CEU program. However, the advantages of participating in this program are that the TCAA short course is already prepared, is preapproved by the AIA, and the registration fees are paid by the Acoustical Society of America.

1:45

4pAAa2. Technical Committee on Architectural Acoustics short course presentation material. Anthony K. Hoover (Cavanaugh Tocci Assoc., 327 F Boston Post Rd., Sudbury, MA 01776)

The Technical Committee on Architectural Acoustics (TCAA) is a Registered Provider in the American Institute of Architects (AIA) Continuing Education System (CES). The TCAA has developed a standardized introductory short course for architects, called Architectural Acoustics. An architect can earn one continuing education unit by attending this short course, if it is presented by a qualified member of TCAA. The course covers topics in sound isolation, mechanical system noise control, and finish treatments. This paper will cover the course material in order to prepare and qualify potential presenters. There will be time at the end of the paper for questions and answers on the course material. In order to qualify as an authorized presenter for this AIA/CES short course, attendance at this workshop and membership in TCAA are required.

2:45–3:00　Break

3:00

4pAAa3. American Institute of Architects/Continuing Education System (AIA/CES) provider registration and reporting requirements. Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Tpke., Vernon, CT 06066)

The Technical Committee on Architectural Acoustics (TCAA) is a Registered Provider in the American Institute of Architects (AIA) Continuing Education System (CES). The TCAA has developed a standardized introductory short course for architects. The TCAA short course is called Architectural Acoustics and attendance at this one-hour long course can earn an architect one continuing education unit (CEU). This paper will cover the administrative requirements of the AIA/CES, to prepare potential presenters. These requirements include the proper handling of paperwork, so that AIA members may receive credit for the course. Also, the manner in which the course is given is dictated by AIA requirements. TCAA membership and attendance at this workshop are required to qualify as an authorized presenter for this AIA/CES short course.

4p THU. PM

Session 4pAAb

Architectural Acoustics: Predictions and Simulation

Paul T. Calamia, Chair

Rensselaer Polytechnic Inst., School of Architecture, 110 8th St., Troy, NY 12180

Chair's Introduction—3:45

Contributed Papers

3:50

4pAAb1. Modifications to the diffusion model for room-acoustic prediction. Yun Jing and Ning Xiang (School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180-3590)

Recently, a diffusion model has drawn attention in room-acoustic predictions. This paper proposes possible modifications to the diffusion model to predict the reverberation times and sound pressure distributions in enclosures. While the original diffusion model [Ollendorff, Acustica **21**, 236–245 (1969); J. Piacut *et al.*, Acustica **83**, 614–621 (1997); Valeau *et al.*, JASA **119**, 1504-1513 (2006)] usually has a good performance for low absorption surfaces, the modified diffusion model, including a new boundary condition and a higher order approximation of a transport equation, yields more satisfactory results for both low and high absorption surfaces. Examples of cubic rooms, flat rooms, and long rooms with various absorption coefficients are presented for comparisons between the modified model, the original model, a geometrical-acoustics model, and several well-established theories in terms of reverberation times and sound pressure level distributions. This paper will discuss improved prediction accuracies by the modified diffusion model.

4:05

4pAAb2. Investigation of voice stage support: A subjective preference test using an auralization system of self-voice. Cheuk Wa Yuen and Paul T. Calamia (Grad. Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, wa@amusical.org)

The human voice plays an integral role in dramatic art. The performance of singers and actors, who perceive their voice through their ears as well as bone conduction, is highly related to the acoustic condition they are in. Due to the proximity of the sound source and the spectral difference in the transmission through the skull as compared to air, a support condition different from that for musical instrumentalists is needed. This paper aims at initiating a standardization of methodology in subjective preference testing for voice stage support in order to collect more data for statistical analysis. A proposal of an acquisition/auralization system for self-voice and a set of subjective test procedures are presented. The subjective evaluation of the system is compared to previous designs reported in the literature, and the implementation is validated. A small playhouse has been measured and auralized using the system described, and subjective-preference tests have been conducted with 13 professionally trained actors. Their preferred stage-acoustic conditions (in relation to locations on stage and head orientations) are reported. The results show potential directions for further investigations and identify the necessary concerns in developing an objective parameter for voice stage support.

4:20

4pAAb3. Validation study of computer modeling of source directivity. Michelle C. Vigeant and Lily M. Wang (Architectural Eng. Prog., Univ. of Nebraska–Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, mvigeant@mail.unomaha.edu)

Previous research has focused on the changes in objective and subjective results when changing source directivity in a computer model, but little work has been done to validate these simulated changes experimentally. An investigation was carried out to compare the measured and simulated room acoustics parameters which result from using a directional sound source, a JBL Eon-10 G2 loudspeaker. Measurements were made in a 100-seat lecture-style classroom, using the sine sweep method, with (a) an omni-directional and (b) the directional loudspeaker oriented in four directions. The measured differences in reverberation time (T30) were minimal across the two source types and four orientations, while significant differences resulted for early decay time (EDT) and clarity index (C80). An ODEON v6.5 model of the classroom was calibrated against the omni-directional results to within two just-noticeable-differences (JNDs) across the three parameters of interest: T30, EDT, and C80. Simulations with the directivity of the JBL loudspeaker were then performed and the results differed by less than two JNDs from the measurements for all source-orientation/receiver combinations. In conclusion, ODEON v6.5 does accurately model the changes in room acoustic parameters which result from using different source directivities. [Work supported by the National Science Foundation.]

4:35

4pAAb4. Modeling room impulse response by incorporating speaker polar response into image source method. Zhixin Chen and Robert Maher (Dept. of Elec. and Computer Eng., Montana State Univ., Bozeman, MT 59717-3780)

Simple computer modeling of impulse responses for small rectangular rooms is typically based on the image source method, which results in an impulse response with very high time resolution. The image source method is easy to implement, but the simulated impulse responses are often a poor match to measured impulse responses because the description of the source is often too idealized to match the real measurement conditions. For example, the basic image source method has often assumed the sound source to be an omni-directional point source for ease of implementation, but a real loudspeaker may include multiple drivers and exhibit an irregular polar response in both the horizontal and vertical directions. In this paper, an improved room impulse response computer modeling technique is developed by incorporating the measured horizontal and vertical polar responses of the speaker into the basic image source method. Results show that compared with the basic image source method, the modeled room impulse response using this method is a better match to the measured room impulse response.

3174   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   3174

4pAAb5. Increasing the direct-to-reverberant ratio in rooms with a constant-beamwidth, wide-bandwidth loudspeaker array. Douglas F. Winker (Acoust. Systems—A Div. of ETS-Lindgren, 415 E. St. Elmo Rd., Austin, TX 78745) and Elmer L. Hixson (Univ. of Texas, Austin, TX 78712)

A constant-beamwidth, wide-bandwidth (CBWB) loudspeaker array was used to improve speech intelligibility in two rooms. The CBWB loudspeaker array maintained a constant beamwidth between 500 Hz to 4 kHz. The array had a narrow vertical beam pattern and a wide horizontal beam pattern. During the study, an increase in the direct-to-reverberant ratio was observed that correlated with an increase in speech intelligibility (measured with the STI method). The correlation between improved speech intelligibly and an increased direct-to-reverberant ratio will be discussed, and the relation between each source's beam pattern and the direct-to-reverberant ratio will be presented. The array was compared to a typical PA source and showed increases in the direct-to-reverberant ratio. Room simulations were conducted with CATT-Acoustic and compared to real-world measurements. The study showed an increase in the direct-to-reverberant ratio in each room, which corresponded to improved speech intelligibility.

4pAAb6. Qualitative analysis of concert hall acoustics with multiple directional sources. Youngmin Kwon, Gary W. Siebein, and Martin A. Gold (Architecture Technol. Res. Ctr., Univ. of Florida, 134 ARCH, P.O. Box 115702, Gainesville, FL 32611, ykwon@acousticdimensions.com)

Following the previous work on quantitative analysis of concert hall acoustics with multiple directional sources (16 directional loudspeakers) [Kwon and Siebein, J. Acoust. Soc. Am. **120**, 3263(A) (2006)], this paper discusses qualitative assessments of concert hall acoustics by means of psycho-acoustical listening tests of room acoustical qualities such as loudness, reverberance, clarity, warmth, spaciousness, envelopment, etc. The music signals binaurally recorded in a performance hall with the same multiple directional source array were evaluated in comparison to those recorded with the single omni-directional source. Results showed that some of the room acoustical qualities of the music signals recorded with multiple directional sources were perceived noticeably different from those recorded with a single omni-directional source. In the main orchestra area at 95% significance level, the music recording with multiple directional sources was perceived more in reverberance and clarity but less in warmth. Some results from these qualitative assessments were, however, found to contradict their quantitative measures.

THURSDAY AFTERNOON, 7 JUNE 2007

SALON A, 1:30 TO 3:20 P.M.

Session 4pEA

Engineering Acoustics: Acoustical Measurements and Methods II

Daniel M. Warren, Chair
Knowles Electronics, 1151 Maplewood Dr., Itasca, IL 60143

Chair's Introduction—1:30

Contributed Papers

4pEA1. 60 MHz multidomain ultrasonic transducer. Igor V. Ostrovskii and Andriy B. Nadtochiy (Dept. of Phys., Univ. of Mississippi, University, MS 38677)

Multidomain periodically poled structure of 80 inversely poled domains of 50-micron-width each was fabricated in a 350-micron-thick 42-degrees-Y-rotated cut of LiTaO3 wafer. A multidomain structure in a ferroelectric plate is characterized by its own domain resonance where acoustic wavelength is equal to the domain period, and ultrasound speed is ∼6 km/s for LiTaO3 [I. V. Ostrovskii and A. B. Nadtochiy, Appl. Phys. Lett. **86**, 222902 (2005)]. The rf impedance of the transducer as a function of frequency demonstrates a sharp resonance-antiresonance behavior at 57.6–60 MHz band (sample LTYR42B5a). The resonance observed is a pure multidomain resonance because crystal length is much longer than domain width. Overall dimensions of this transducer are $4 \times 4 \times 0.35$ mm^3, and irradiating surface area is 4×0.35 mm^2. Two such transducers were used to model a delay line with water as a medium of propagation. A 1-μs burst with 59 MHz central frequency was delayed for 4.3 μs in the water. In conclusion, we present a new high frequency transducer for various applications in non-destructive testing, medicine, and biology [Patent Pending]. Experimental measurements and theoretical calculations are in good agreement.

4pEA2. Head and torso simulators. Gunnar Rasmussen (Skovlytoften 33, 2840 Holte, Denmark, gr@gras.dk)

The KEMAR was developed by Knowles Electronics and introduced in 1972, thoroughly described by Mahlon D. Burkhard, and standardized in ANSI S3.361ASA58-1985. The KEMAR was provided with ears typical of American and European females as well as Japanese males and females. Larger ears more typical of American and European males have been developed with well-defined hardness of 0055 and 0035 shore. The ears are readily exchanged with the ITU P57 Type 3.3 ear. The KEMAR is designed to offer lifelike reproducible test conditions and is therefore based on average anthropometric dimensions of an adult human, while other HATS are designed using simplified or average geometric convenient dimensions. The KEMAR may be provided with an artificial mouth system with well-defined characteristics. The KEMAR also offers well-defined positioning of microphones for linearization as well as positioning systems for handheld communication systems. Measurements show good agreement with early data published by M. D. Burkhard and R. M. Sachs ["Anthropometric Manikin for Acoustic Research," JASA **58**, 214–222 (1975)]. Conclusion: Acoustical measurements carried out since 1972 can still be compared and carried out on a system, which has not changed acoustically except for increased versatility and improved acoustic documentation.

4p THU. PM

4pEA3. An etude in hearing-aid directivity measurement and benchmarking. Matthew J. Green and Thomas Burns (Starkey Labs, 6600 Washington Ave. S., Eden Prairie, MN 55344)

In this study, a three-dimensional, automated scanning system is used to measure the directional performance of hearing aids mounted on a test manikin. A diffuse field is simulated in an anechoic environment by sequentially positioning a loudspeaker along the surface of a 1-m-radius sphere circumscribing the manikin and device under test. Particular attention is given to the type of stimulus, the spatial resolution of loudspeaker positions, and postprocessing of the impulse responses; each affects the measurement precision of the Directivity Index and the 3D polar responses. The purpose of this research is to describe the recipe that yielded minimal measurement bias and variance for ITEs, BTEs, and OTEs of varying directivity.

4pEA4. Methodology of planar unsteady flow measurements in an excised larynx. Shanmugam Murugappan, Sid Khosla, Ephraim Gutmark, and Raghava LakhamRaju (Dept. of Otolaryngol.-HNS, Univ. of Cincinnati, 231 Albert Sabin Way, #MSB 6308, Cincinnati, OH 45267)

A methodology to obtain phase locked planar unsteady flow of the near field of supraglottal flow in an excised larynx using particle image velocimetry (PIV) is presented. Different steps involved in obtaining 2D flow field at different instances of the EGG or microphone signal and the structural motion of the vocal fold are discussed. Real time signal processing of the sensor signal coupled with the synchronization of the PIV flow field and high speed imaging of the vocal fold is used to evaluate the relationship between acoustics-structure and flow. Issues that can affect the accuracy of the flow field measurements and the reconstruction of the flow patterns at different instances during the EGG/microphone cycle were identified and suitable measures taken to resolve them are discussed. Some of these include simulating a symmetric vocal fold motion, maintaining a constant phonation frequency over the entire experimental run, preventing any spurious vectors that could arise due to reflection from the moist tissue surface or poor spatial or temporal resolution. Examples are also shown where the effect of these factors impact both the raw seeded image and correlation PIV algorithm used to extract the flow field vectors.

4pEA5. Flow induced aeroacoustic oscillations in a tube with a region of abrupt cross-sectional area increase. Donald Bliss and Noah Sakimura (Mech. Eng., Duke Univ., Durham, NC 27708)

The phenomenon of aeroacoustic oscillations induced by flow through a circular tube with an annular cavity region formed by an abrupt area increase is studied. At the beginning of the area increase, the flow separates and produces a shear layer that subsequently impinges on the downstream end of this region. The shear layer is a medium for propagation of convective vorticity waves. In the surrounding cavity, axial acoustic wave propagation can occur. The central flow region is governed by the convective wave equation, and the simple wave equation applies in the cavity. Coupled axial modes are identifiable as being primarily convective or primarily acoustic. The oscillation mechanism involves the interaction of upstream acoustic modes and downstream convective modes, coupled together by boundary conditions, and driven by mass addition and removal at the rear of the annular cavity caused by shear layer impingement on downstream edge. The oscillation process is analyzed using a cross-sectional averaging method for quasi one-dimensional waves to determine system wavenumbers and eigenmodes. Oscillation frequencies are determined by the condition of maximum energy extraction from the mean flow, arising from unstable convective waves, and by a compatible shear layer phase condition. A comparison with simple experiments is presented.

4pEA6. Chaotic state in an electrodynamic loudspeaker controlled by gas pressure. Ivan Djurek, Antonio Petosic (Faculty of Elec. Eng. and Computing, Dept. of Electroacoustics, Unska 3, HR-10000, Zagreb, Croatia), and Danijel Djurek (AVAC, Kesten brijeg 5, Remete, HR-10000 Zagreb, Croatia)

An electrodynamic loudspeaker with resonant frequency $f_0 = 47.1$ Hz has been driven in the far nonlinear regime, and f_0 increases with increasing driving ac current I_0. Landau cutoff of the vibration amplitude appears at frequency $f_c > f_0$, which is followed by the doubling of driving period $1/f$ and appearance of harmonic sequences at $\frac{1}{2}nf$, $\frac{1}{4}nf$, $\frac{3}{4}nf$, ... By further increase of current the white noise spectrum appears, which is characteristic of the chaotic state. Electrodynamic loudspeaker is represented by an ordinary differential equation of motion describing an anharmonic forced oscillator, and it is possible to achieve an independent control of chaotic state by the gas pressure, since real R_S and imaginary part X_S of the gas acoustic impedance affect, respectively, friction and inertial terms in the equation. The used gas atmospheres ($0.01 < p < 1$ bar) were H$_2$, D$_2$, He$_4$, Ne, Ar, CO$_2$, SF$_6$, and air. The cutoff frequency f_c depends on the gas pressure and it was plotted against (p, I_0), which in turn defines the surface in three-dimensional diagram. Values of parameters (p, I_0) triggering chaotic state were chosen above this surface, and universality of some laws describing such state was tested against the gas density.

4pEA7. Ultrasonic resonance spectroscopy for cryogenic materials. Petre Petculescu, Remus Zagan (Ovidius Univ., 124 Mamaia Blvd., Constanta 900572, Romania), and Andi Petculescu (Univ. of Louisiana-Lafayette, Lafayette, LA 70504)

Cryogenic steels are needed in the construction of industrial installations able to withstand low temperatures. This poster presents results on the use of an ultrasonic resonant technique to determine the composition of the cryogenic stainless steel 10TiNiCr180 with an austenitic structure. The investigations were done both in the base material and in the weld, with and without thermal treatment. Upon exposing the sample to two thermal treatments (welding-hardening and welding-hardening-annealing), metallographic analysis showed the presence of chromium carbide in the base material and the weld. The ultrasonic power spectrum shows a prevalent austenitic structure in the base material and a compound austenite-ferrite-chromium carbide structure in the weld. The grain size determined from the spectral peak ratio is compared with the one obtained by metallographic analysis. The measurements presented here were obtained in direct contact with the pulse-echo technique, with excitation centered at 5 MHz.

Session 4pMU

Musical Acoustics: "Myths and Mysteries" in Musical Acoustics

Evan B. Davis, Cochair
Boeing Commercial Airplanes, 8556 Burke Ave., N., Seattle, WA 98103

Thomas R. Moore, Cochair
Rollins College, Dept. of Physics, Winter Park, FL 32789

Chair's Introduction—2:00

Invited Papers

2:05

4pMU1. Myth-busting brass instrument design. Richard A. Smith (Smith-Watkins Brass, Richard Smith (M I) Ltd., Cornbrough, Sheriff Hutton, York, YO60 6RU, UK)

Richard Smith has designed and made instruments for professional brass players for over 35 years. With a background as scientist and musician, he has a unique view of the subject in which his myth-busting has been a dominant feature. The challenge now is to transfer this knowledge to music educators and players. It can be shown that any instrument is a compromise of musical qualities, making the quest for the perfect instrument pointless and a distraction from the real business of making music. Resistance is probably the most difficult of these musical qualities to interpret from player responses. It will be demonstrated how the language of fluid dynamics has been wrongly adopted by players, educators, and manufacturers to give pseudo-scientific authority to their teaching and product promotion. The development of a blindfold testing technique by Smith was crucial to (a) the study of materials used in instrument construction and (b) a breakthrough in efficient instrument design. Current work with Thomas Moore (Rollins College, USA) and David Howard (University of York, UK) investigates the mystery of hi-note playing on the trumpet where the most skilled players find distinct slots above 2 kHz in, according to acoustical textbooks, a resonance free zone.

2:25

4pMU2. It's all in the bore![a] −Is that true? Aren't there other influences on wind instrument sound and response? Wilfried Kausel (Inst. of Musical Acoust., Univ. of Music and Performing Arts, Anton von Webernplatz 1, A-1030 Vienna, Austria, kausel@mdw.ac.at)

A review of the current state of knowledge concerning the influence of wall vibrations on the sound of brass wind instruments, flue pipes, and woodwind instruments is given. The question of whether this influence is strong enough to be objectively observable is still a controversial issue. While instrument makers and musicians make a strong claim that wall thickness, material, and conditioning are crucial factors for sound quality and response of wind instruments, acousticians tended—at least in the past—to attribute all acoustical characteristics to the actual bore profile. Recently this seems to be changing. First, there is growing experimental evidence favoring the hypothesis that wall vibrations do matter; second, theoretical models are starting to be developed which deal with the interaction between the air column sound field and its oscillating boundary conditions. Another suggested theory is that wall vibrations are propagated back to the mouthpiece interacting with the player's lips affecting the delicate and strongly nonlinear mechanism of sound generation. Even the question of whether vibrating bells can radiate a significant amount of sound does not seem to have reached a final answer. New experiments have been conducted to help answer these open questions. [a]R. A. Smith, "It's all in the bore!," International Trumpet Guild Journal, **12**(4) (1988).

2:45

4pMU3. Vibroacoustics of musical wind instrument: Analysis of some pathological and nonpathological behaviors. Francois Gautier, Guillaume Nief, Joel Gilbert, and Jean-Pierre Dalmont (Laboratoire d'Acoustique de l'Universite du Maine, UMR CNRS 6613, Av. O. Messiaen, 72085 Le Mans cedex 9, France)

Wall vibration of a musical wind instrument can be clearly noticed and measured in playing configuration. However, the problem of quantifying its effect on the emitted sound remains a subject of debate. Wall vibrations are generated by two mechanisms: (1) a mechanical source corresponding to the impacts of a reed (or musicians lips) on the mouthpiece and (2) an acoustical source due to the internal sound field. In this paper, we present an investigation of mechanism (2) using an experimental approach and a theoretical model of a generic simplified instrument (cylindrical vibrating shell, with a slightly distorted circular cross section). Analysis leads to the conclusion that, overall, vibroacoustic couplings are very small and do not induce any audible contribution. However, the wall vibration can play a significant role for some particular choices of material and geometry, which lead to coincidences between structural and acoustical modes. In these configurations, model and experiments show that the input acoustic impedance is clearly perturbed by wall vibrations. Using a blowing machine, it is shown that these perturbations can induce some timbre changes or some unstable oscillations regimes. Analysis of such pathological behaviors will be presented.

4p THU. PM

4pMU4. On the need for musical instrument strings to dynamically stretch. Evan B. Davis (Boeing Commercial Airplanes, 8556 Burke Ave N., Seattle WA 98103)

Given clamped or fixed boundary conditions a string must stretch or elongate to vibrate. Many have analyzed the effect of string stretch under the fixed boundary condition assumption and the related nonlinear motion of the string. However, if the boundary of the string at the soundboard was fixed the soundboard would not move. A rigid soundboard is not a very useful design for the production of music. It will be shown that there are geometries which allow soundboard motion without string stretch well as geometries that require string stretch for soundboard motion.

Contributed Papers

<table>
<tr><td>3:25</td><td>3:40</td></tr>
</table>

4pMU5. To play it is to improve it: An analysis of the "playing-in" phenomenon of stringed instruments. Andrew A. Piacsek (Dept. of Phys., Central Washington Univ., Ellensburg, WA 98926-7422)

There is a growing interest in resolving the question of whether, and how, the act of playing a stringed instrument (in the violin and guitar families) improves its sound and playability over time. Although generally accepted by musicians, the causal connection is difficult to test due to the very large number of potentially relevant parameters, both physical and psychological. This presentation describes various historical references, interviews with musicians, and recent research on the subject of the "playing-in" phenomenon. Results will be presented from new experiments on the response of spruce panels subject to mechanical stimulus.

4pMU6. How can bell vibrations affect the sound of brass instruments? Thomas Moore, Vishal Jiawon, and Daniel Zietlow (Rollins College, Winter Park, FL 32789, tmoore@rollins.edu)

It has recently been shown that the vibrations of the bell of a trumpet affect the spectra of the sound produced during play. The manner in which these vibrations produce this effect is still under investigation, but two possibilities have been proposed: (1) the vibrations of the bell feedback to the lips and change their motion, and (2) the bell vibrations change the radiation impedance of the instrument. The former explanation appears to be the most plausible; however, we will present results of recent experiments that indicate at least some of the effect is due to a change in radiation impedance.

THURSDAY AFTERNOON, 7 JUNE 2007 CANYON ROOM A, 1:50 TO 5:00 P.M.

Session 4pNS

Noise, Engineering Acoustics, and Signal Processing in Acoustics: Recent Advances in Active Noise Control II

Scott D. Sommerfeldt, Cochair

Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602-4673

Kenneth A. Cunefare, Cochair

Georgia Inst. of Technology, School of Mechanical Engineering, 113 MRDC II, Love Building Atlanta, GA 30332-0405

Invited Papers

4pNS1. Advanced digitally adjustable analog feedback control system and its usage in active noise insulation. Marko Antila, Hannu Nykanen, Jari Kataja, and Velipekka Mellin (VTT Tech. Res. Ctr. of Finland, POB 1300, FIN-33101 Tampere, Finland)

Feedback control for the active noise cancellation can be done with analog or digital circuitry. Although digital approach is steadily gaining ground, analog control still has its uses. Short control latency is the feature of an analog controller—it has the shortest possible delay. Its drawback is the limited flexibility for adjustable or adaptive solutions. To overcome this, digitally adjustable analog controllers can be used. A potential implementation for such controller uses field programmable analog array (FPAA) technology. With FPAAs both the filter parameters and structure are digitally adjustable, even during the operation. A development environment around the FPAA controller has been devised. It consists of the measurement system, an optimizer that calculates the optimum controller response, programming software, and actual FPAA hardware with its embedded software. The complete system is packaged tightly to a compact form. It can be used for evaluation purposes in acoustics laboratory conditions. Such a system has been evaluated in active noise insulation application. It has proven to operate as simulated and optimized also in real-world tests. The future development of the system includes further miniaturization, controller networking, and possible integration inside the sound insulation system.

4pNS2. Rapidly converging adaptive state-space-based multichannel active noise control algorithm for reduction of broadband noise. Arthur P. Berkhoff (TNO Sci. and Industry, P.O. Box 155, 2600AD Delft, The Netherlands, arthur.berkhoff@tno.nl) and Johan M. Wesselink (Univ. of Twente, 7500AE Enschede, The Netherlands)

Rapidly changing spectra may lead to performance limitations in adaptive systems for broadband active noise control, especially in multichannel systems. This paper presents techniques to address the negative consequences of two main causes. First, the dynamics of the transfer paths between the noise control sources and the error microphones is compensated for by using a regularized state-space based adaptive filtered-error scheme. Second, for the reference signals a modified adaptive scheme is used, taking into account the nonwhiteness of these signals as well as the correlation between the individual signals. Examples are given for simulated data and for real-time implementations. [This work was partly supported by EC Contract 501084 (InMAR).]

4pNS3. Optimization of control source locations in an active noise control application of axial cooling fans using genetic algorithms. Connor R. Duke, Scott D. Sommerfeldt, and Kent L. Gee (Brigham Young Univ., N-281A ESE Provo, UT 84602, crd26@byu.edu)

By placing control sources in the near field, global attenuation of an axial cooling fan's blade passage frequency and harmonics can be achieved using active noise control. The optimal position of control sources when modeling a fan, whose characteristics differ significantly from a single monopole source or when modeling more than one fan, will be discussed. The number of control sources, as well as position of control sources, are optimized for specific primary source configurations using genetic algorithms. Source strength of control sources are calculated analytically to minimize radiated sound power. The results of optimization using different forms of genetic algorithms will be presented and compared. The results obtained from the genetic algorithms will be compared to experimental results using multiple sources.

4pNS4. Noise annoyance reduction using active control. Gerard Mangiante and Georges Canevet (Laboratoire de Mecanique et d'Acoustique, CNRS, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France, mangiante@lma.cnrs-mrs.fr)

In two former studies [G. Canevet and G. Mangiante ''Psychoacoustic Assessment of Active Noise Control,'' Active 2004 (2004), and G. Mangiante and G. Canevet ''Active Reduction of Environmental Noise: A Sound Quality Evaluation,'' Internoise 2006 (2006)], the psychoacoustical effects of active noise control were described. One of these effects is the reduction of noise annoyance. This contribution aims at describing this annoyance reduction using the following indicators: loudness, sharpness, unbiased annoyance, and acoustic comfort. These indicators were calculated with a psychoacoustic-analysis software in which the following models were implemented: Moore and co-workers (1997), Zwicker and Fastl (1999), and Glasberg and Moore (2002). Using this software, various noises, selected from our everyday environment, were tested to investigate the effects of the following parameters: the width and the center frequency of the controlled band, the maximum attenuation in this band, and the sound pressure level of the noise before active control. The results showed that, in all cases, acoustic comfort is positively affected by active control. The annoyance reduction is greatly dependent of the shape of the power spectrum of the noise to be reduced. The best results are obtained for broadband noises with an energy contribution concentrated at low frequencies.

4pNS5. Eigenvalue equalization applied to the active minimization of energy density in a mock helicopter cabin. Stephan P. Lovstedt (Dept. of Phys. and Astron., Brigham Young Univ., N283 ESC Provo, UT 84602, s_lovstedt@yahoo.com), Jared Thomas, Scott D. Sommerfeldt, and Jonathan Blotter (Brigham Young Univ., Provo, UT 84602)

A number of applications in active noise control require the ability to control and track multiple frequencies. If a standard filtered-x algorithm is used, the system must be designed to be stable for the slowest converging frequency anticipated, thereby leading to reduced overall performance of the system. Previous work has focused on overcoming this through development of a method that equalizes the eigenvalues of the system over the operating frequency range, leading to more uniform performance. The current work has built on the previous work to extend the method for implementation in systems that control the acoustic energy density. Minimizing energy density has been shown to have favorable performance characteristics when used for controlling enclosed acoustic fields. Thus, combining the approach of equalizing the system eigenvalues with energy density control leads to a system that incorporates the advantages of both methods. The control approach is demonstrated through implementation in a mock helicopter cabin, to demonstrate the favorable convergence characteristics, along with the global control of the field.

3:30–3:45   Break

4p THU. PM

Contributed Papers

4pNS6. Active noise reduction communication earplug for helicopter crew. Jason A. Solbeck, Matthew Maher, Toby Deitrich (Sound Innovations, Inc., 55 Railroad Row, White River Junction, VT 05001), Laura R. Ray, and Robert D. Collier (Dartmouth College, Hanover, NH 03766)

A custom designed shallow insert communication earplug for helicopter crews consists of a system based on a patented hybrid ANR feedback-feedforward algorithm [Ray *et al.*, J. Acoust. Soc. Am. **120**(4), 2026–2038 (2006)]. The hybrid system provides the benefits of feedback ANR while extending the bandwidth and magnitude of total ANR performance with feedforward ANR, which improves attenuation of tonals. The development includes optimized algorithms, a custom earplug, and a miniaturized battery powered ANR module, weighing approximately 250 g. The hybrid ANR earplug requires two miniature microphones, one inside and one external to the earplug and an internal speaker to deliver the cancellation signal. The software automatically adapts the ANR algorithm to the transfer function characteristics of the system to both improve performance and accommodate individual human fitting differences. The testing protocol and results are based primarily on measurements with a HEAD Acoustics manikin in the Sound Innovations sound room [Duncan *et al.*, Proceedings of Internoise 2006, December 2006]. MIRE testing is planned for the Spring of 2007. Results to-date demonstrate comparable passive attenuation to a commercially available communication earplug and an added average active attenuation of approximately 6 dB for UH-60 helicopter noise. [Work supported by U.S. Army.]

4pNS7. NASA-sponsored active fan noise control research. Joe Posey and Michael Jones (NASA Langley Res. Ctr., Hampton, VA 23681, joe.w.posey@nasa.gov)

The National Aeronautics and Space Administration (NASA) has espoused the vision of developing technology capable of keeping all objectionable aircraft noise within airport boundaries. In order to achieve such an aggressive goal, new aircraft will have to be designed employing quiet propulsion, quiet lift, and quiet drag. Research activities are continuing on all these fronts. Dominant sources of propulsion noise to date on jet-powered transports have been jet noise and fan noise. As bypass ratios of aircraft turbofan engines have increased to improve fuel efficiency, jet velocities (and therefore jet noise) have decreased, making fan noise the larger contributor in many instances. In NASAs Advanced Subsonic Technology (AST) Program, which ran from 1994 until 2001, several active control strategies for tonal fan noise were investigated. These included rings of wall-mounted actuators, active impedance control, stator-mounted actuators, and hybrid active/passive concepts. The most promising approach seems to be stator-mounted actuators to provide control sources effectively co-located with the rotor/stator interaction source. The major remaining issue for active control of fan tones is the availability of high-powered, compact, efficient actuators for stator mounting. Control of broadband fan noise has been addressed, but much work remains to be done.

4pNS8. Determination of optimal near-field error sensor locations for active control of cooling fan noise using spherical harmonic expansions. Benjamin M. Shafer, Kent L. Gee, Scott D. Sommerfeldt, and J. Isaac Fjeldsted (Dept. of Phys. and Astron., N283 ESC, Provo, UT 84602, otishobbes@gmail.com)

Recent developments in the active control of cooling fan noise have used near-field error sensors to drive the pressure to zero [K. L. Gee and S. D. Sommerfeldt, J. Acoust. Soc. Am. **115**, 228–236 (2004)]. Theoretical mapping of near-field pressure during minimization of sound power re-

veals the location of pressure nulls that can be used to optimize the location of the error sensors. To this point, the locations of error sensors have been determined by modeling both the fan and the control loudspeakers as point monopoles. However, noise from an axial fan has monopole as well as multipole characteristics. The multipole characteristics of the fan can be obtained using a procedure based on the work of Martin and Roure [T. Martin and A. Roure, J. Sound Vib. **201**, 577–593 (1997)]. Pressure values are obtained over a hemisphere in the far field of a primary source and the contributions from multipoles up to the second order, centered at the primary source, may be calculated using spherical harmonic expansions. The source information is then used in the aforementioned theoretical near-field calculation of pressure. Error sensors are positioned using this calculation and the global attenuation is compared to that of previous experiments.

4pNS9. Near field placement of error sensors in an active noise control application of axial cooling fans using flow visualization techniques. Connor R. Duke, Scott L. Thomson, Scott D. Sommerfeldt, Kent L. Gee Gee (Brigham Young Univ., N-281A ESE Provo, UT 84602, crd26@byu.edu), Cole L. Duke, and David W. Krueger (Brigham Young Univ., Provo, UT 84602)

The use of error sensors in the near field of an axial fan can be used to achieve global attenuation of the blade passage frequency and its harmonics. The pressure field produced by the minimized sound power radiation of the fan and control source configuration dictates possible locations of the error sensors by creating pressure nulls in the near field. By minimizing pressure at these locations, the minimized sound power field can be reproduced. Near field locations can be further investigated using qualitative flow visualization techniques including smoke visualization and quantitative particle image velocimetry (PIV). Analysis of the flow field will help to understand the presence of turbulent flow at the error sensors which can be a cause of a decreased signal to noise ratio. Optimal locations as well as mounting techniques for the error sensors will be discussed based on the flow field analysis. The effects of active noise control on the flow field produced by the fan will also be discussed.

4pNS10. Further developments in the active control of free-field sound using near-field energy-based error signals. Ryan T. Chester and Timothy W. Leishman (Phys. Dept., Brigham Young Univ., N283 ESC, Provo UT 84602)

Practical efforts to actively control sound often require error sensors located in the acoustic or geometric near field of sound sources. Unfortunately, when using conventional acoustic pressure sensors, control performance often becomes very sensitive to sensor location. Optimal sensor positions are difficult to ascertain or maintain with changing system conditions. Alternate types of error sensors often lead to fewer position-dependent complications, producing global sound field control with greater ease and consistency. The distinct spatial and spectral uniformities of potential, kinetic, and total energy density for near-field error signal measurements will be discussed for the sound control of sources located in a free field. Numerical results will be compared to experimentally measured results.

Session 4pPA

Physical Acoustics: Bubbles, Modes, and Scattering

Mark S. Wochner, Chair

Univ. of Texas, Applied Research Lab., 10000 Burnet Rd., Austin, TX 78758

Chair's Introduction—1:40

Contributed Papers

1:45

4pPA1. Sonoluminescence from transient cavitation at high pressures in water and acetone. D. Felipe Gaitan and Ross A. Tessien (Impulse Devices, Inc. 13366 Grass Valley Ave., Grass Valley, CA 95945)

In highly degassed, clean liquids, transient acoustic cavitation can be triggered by fast neutrons, a phenomenon that has been known since ~1960's. The kinetic energy acquired by the nuclei allows it to ionize a small volume (~100 nm dia.) in the liquid, creating a vapor cavity that would normally last a couple of microseconds. If the acoustic amplitude and phase are right, this cavity expands by several orders of magnitude (~500 to 1500 microns dia.) and then collapses, emitting a short flash of light (1 to 40 nsec). The bubble continues to expand and collapse for several hundred cycles, eventually evolving into a larger (2 to 6 mm dia.) bubble cloud which lasts several milliseconds, depending on the conditions. The time duration of the light pulses is longer and their amplitude larger than those of single bubble sonoluminescense in water (~100 to 300 psec, 10^5 photons/flash). The amplitude and time evolution of the light flashes has been analyzed as a function of the driving conditions and compared with computer simulations in an effort to infer the maximum plasma temperatures and densities, and perhaps the presence of shock waves, in the cavities.

2:00

4pPA2. Sonoluminescence: Theoretical explanation and comparison to experiment pertaining to the transduction of sound into radiated light. Harvey Woodsum (Sonetech Corp., 10 Commerce Park North, Unit 1, Bedford, NH 03110)

The subject of sonoluminescence, i.e., the transduction of sound into light via the intermediate step of cavitation bubble collapse, is of general interest and has sparked a great deal of published research. The present work considers the integration of bubble dynamics with the nonlinear generation of light via the Euler-Heisenberg theory [W. Heisenberg and H. Euler, Z. Phys. **98**, 714 (1936)], in relation to Schwingers' conjecture that bubble cavitation is ''squeezing light from the quantum vacuum'' via the Casimir effect [Schwinger, Proc. Natl. Acad. Sci., USA **90**, 4505–4507, (1993)]. Preliminary results of applying this model appear to show good agreement with both radiated spectral shape and efficiency of sonoluminescent light production, as presented by the relevant literature.

2:15

4pPA3. Effects of liquid compressibility on coupled radial and translational motion of a bubble. Todd A. Hay, Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

In shock wave lithotripsy, explosive bubble growth is followed by a collapse phase during which bubble wall velocities can approach the speed of sound in the liquid. If the bubble is simultaneously in translation, the velocity of translation can also exhibit a very large spike during rebound. The velocity spike corresponds to conservation of translational momentum as the bubble radius, and thus entrained mass, become very small. It can then be important to include effects of compressibility on both the radial and translational motion. This presentation discusses corrections to the coupled equations for radial and translational motion when compressibility, and therefore acoustic radiation, are taken into account. For pure radial motion, compressibility introduces a radiation term proportional to the third time derivative of bubble volume, and inversely proportional to sound speed. For pure dipole motion, linear theory predicts a radiation force proportional to the second time derivative of the translational velocity, also inversely proportional to sound speed [Morse and Ingard, *Theoretical Acoustics* (1968)]. Numerical simulations will be presented that illustrate the magnitudes of compressibility effects on solutions of the coupled equations for radial and translational motion. [Work supported by the NSF Graduate Research Fellowship.]

2:30

4pPA4. Excitation of shape oscillation of bubble(s) in water under ultrasonic vibration. Ichiro Ueno, Tatsunori Kojo, and Shuhei Katase (Tokyo Univ. of Sci., 2641 Yamazaki, Noda, Chiba 278-8510, Japan)

Behavior of a single or multiple bubbles of O(1 mm) in diameter under ultrasonic vibration was investigated experimentally. Excitations of surface wave and shape oscillation with distinct mode number were especially focused. Interaction between/among oscillating bubbles in a row rising in water was also introduced by considering bubbles volume variation during the oscillation. Through a series of the present experiment, variation of distinct mode number for shape oscillation as a function of the bubble radius is indicated.

2:45

4pPA5. Excitation of a vertical liquid capillary jet waveguide by modulated ultrasonic radiation pressure. J. B. Lonzaga, D. B. Thiessen, and P. L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

The excitation of a liquid capillary jet issuing from a nozzle is investigated using internally applied modulated ultrasonic radiation pressure. The transducer used here is more efficient than one used in prior studies [J. B. Lonzaga *et. al.*, J. Acoust. Soc. Am. **116**, 2598 (2004)] and is suitable for nozzle velocities as low as 25 cm/s. At low velocities, the liquid jet is significantly tapered. As a consequence of the taper, the acoustic cutoff frequency increases with increasing distance from the nozzle. For ultrasound propagation down the jet from the nozzle, finite-element calculations of the radiation pressure show that the radial stress at the cutoff location is significantly larger than at any other region of the jet. Two distinct capillary modes exist and are conjectured to be excited near the cutoff location by modulated radiation pressure: one traveling upward and the other traveling downward. In a certain range of modulation frequencies, the latter is an exponentially growing mode and leads to the forced disintegration of the jet into liquid drops. Excitation of the growing mode at the nozzle is achieved for low-speed jets by using carrier frequencies close to or below the nozzle cutoff frequency. [Work supported by NASA.]

4p THU. PM

4pPA6. Scholte-Stoneley wave generation by means of diffraction on a corrugated surface: Influence of corrugation shape. Nico F. Declercq (Woodruff School of Mech. Eng., GATech, 801 Ferst Dr., Atlanta, GA 30332-0405, nico.declercq@me.gatech.edu and GT-Lorraine, 2 r. Marconi, 57070 Metz, France) and Bart Sarens (Katholieke Universiteit Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium)

Because of the growing number of applications of phononic crystals and other periodic structures, there is a renewed and growing interest in understanding the interaction of ultrasound with periodically corrugated surfaces. A theoretical investigation is presented of the transformation of ultrasound incident from the solid side onto a solid-liquid periodically corrugated interface. It is shown that it is possible to tailor the shape of a corrugated surface with given periodicity such that there is a more significant amount of energy transformed into Scholte-Stoneley waves than if pure saw-tooth or sine-shaped surfaces were used. This permits the fabrication of periodic structures that can be patched on or engraved in body parts of a construction and enables efficient generation of Scholte-Stoneley waves. The study is performed for incident homogeneous plane waves as well as for bounded beams. Incident longitudinal waves are studied and also incident shear waves.

4pPA7. Nonlinear shear wave beams. Mark S. Wochner, Mark F. Hamilton, and Evgenia A. Zabolotskaya (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, mwochner@arlut.utexas.edu)

The theoretical framework developed previously for plane nonlinear shear waves [Zabolotskaya et al., J. Acoust. Soc. Am. **116**, 2807–2813 (2004)] has been extended to include diffracting beams. Two coupled KZK-type equations with cubic nonlinearity are obtained for the two transverse components of the particle velocity vector. The coupled nonlinear equations are solved numerically in the time domain using a nonoscillatory numerical scheme. Both linearly and elliptically polarized beams are considered. For circular polarization the equations decouple and become linear. Waveforms along the beam axis are presented that illustrate shock formation and coupling of the two velocity components for different polarizations. Beam patterns are presented showing the angular dependence of the nonlinearly generated harmonics. Because the evolution equations are derived using the parabolic approximation for narrow beams, the predicted beam patterns are axially symmetric for all polarizations whenever the amplitude and phase of the source are axially symmetric. Comparisons are made with the corresponding results obtained from the quadratically nonlinear KZK equation for sound beams in fluids. [Work supported by NIH and ARL IR&D.]

4pPA8. Leaky Rayleigh waves propagating around the corner of a thick solid plate: Finite element analysis. Ebrahim Lamkanfi, Nico F. Declercq (Woodruff School of Mech. Eng., GATech, 801 Ferst Dr., Atlanta, GA 30332-0405, nico.declercq@me.gatech.edu and GT-Lorraine, 2 r. Marconi, 57070 Metz, France), Wim Van Paepegem, and Joris Degrieck (Ghent Univ., Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium)

Finite element simulations reveal the experimentally indicated fact [Declercq et al, J. Appl. Phys. **96**(10), 5836–5840 (2004)] that leaky Rayleigh waves propagating along the horizontal surface of a thick fluid-loaded solid plate are transmitted around the corner of the solid plate. The experiments indicate that leaky Rayleigh waves are generated around the corner of the plate, but the experimental evidence is not fully conclusive whether the effect is caused merely by incident Rayleigh waves on the upper surface or by scattering effects when the incident beam interacts with the corner. The current study first investigates the reported experiments by means of the finite element method. Then the model is applied to study leaky Rayleigh waves separately from the incident and reflected

bounded beams. It is shown that the Rayleigh waves themselves are the physical origin of the transmission of Rayleigh waves and not the scattering effects caused by the incident bounded beam interacting with the edge of the plate.

4pPA9. Non-contact and local characterization of thin plates and cylindrical shells using the zero-group velocities (ZGV) Lamb modes. Dominique Clorennec, Claire Prada, Daniel Royer (Lab. Ondes et Acoustique, ESPCI, Université Paris 7 Denis Diderot, UMR CNRS 7587, 10 rue Vauquelin, 75231 Paris cedex 05, France), and Todd Murray (Boston Univ., Boston, MA 02215)

The group velocity of the symmetrical S_1 Lamb mode vanishes if the Poisson's ratio v is smaller than 0.45. At this ZGV point, the frequency $\times$thickness product undergoes a minimum and two modes (S_1 and S_{2b}) propagate with opposite phase velocities. In our experiments, this ZGV mode is generated by a 10-ns laser pulse and detected at the same point using an optical interferometer. The spectrum of the signal reveals an intense narrow peak corresponding to the S_1-ZGV Lamb mode resonance. A similar phenomena occurs for the A_2 Lamb mode if $v<0.31$. The minima of the frequency$\times$thickness product of the S_1 and A_2 modes only depend on v. The ratio of the resonance frequencies of these two modes is independent of the thickness. Thus, this experimental configuration allows a local and absolute measurement of v. Knowing h, the measurement of these resonance frequencies provides the bulk waves velocities with a high accuracy. This non-contact method was tested on different plates and shells. The same behaviour is observed in a cubic silicon crystal. Moreover in such anisotropic plates, we show that the resonance frequencies can be selected using a line laser source. These frequencies are well predicted by the theoretical dispersion curves.

4pPA10. Periodic orbit techniques for mode shape approximation. Christopher J. Ham and M. C. M. Wright (Inst. of Sound and Vib. Res., Univ. of Southampton, University Rd., Southampton, UK SO17 1BJ, ch@isvr.soton.ac.uk)

At high frequency, numerical methods for calculating mode shape functions in a waveguide are expensive. The effect of modifying the shape of a waveguide in an industrial design context may only be calculated on a case-by-case basis if numerical methods are used. Periodic orbits are closed, repeating ray paths. A significant body of research in quantum physics indicates that periodic orbits provide fundamental insight into mode theory, which can be applied to acoustics. A method to approximate mode shape functions around short periodic orbits proposed by Vergini [J. Phys. A: Math. Gen. **33**, 4709–4716, (2000)] and Babic and Buldryrev ["Short-Wavelength Diffraction Theory" (Springer, Berlin) (1991)] is implemented in this paper. In particular, mode shape functions in nonintegrable waveguides are approximated. This method links the geometry of the waveguide to the mode shape functions. The technique is extended to waveguides that have partly chaotic ray dynamics and is applied to the oval and eccentric annulus waveguides. The approximate mode shape functions from these two cases are compared with numerical results.

4pPA11. Wigner distribution of an ultrasonic transducer beam pattern through a liquid-solid interface. Goutam Ghoshal and Jospeh A. Turner (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, W317.4 Nebraska Hall, Lincoln, NE 68588-0526)

Diffuse ultrasonic backscatter techniques are useful for probing heterogeneous materials to extract microstructural parameters and detect flaws that cannot be detected by conventional ultrasonic techniques. Such experiments, usually done using a modified pulse-echo technique, utilize the spatial variance of the signals as a primary measure of microstructure. Quantitative ultrasonic scattering models include components of both transducer beams as well as microstructural scattering information. Of particular interest for interpretation of many experiments is the propaga-

tion through a liquid-solid interface. Here, the Wigner distribution of the beam pattern of an ultrasonic transducer through a liquid-solid interface is used in conjunction with the Bethe-Salpeter equation to model this scattering problem. The Wigner distribution represents a distribution in space and time of the spectral energy density as a function of wave vector and frequency. The general scattering model to be discussed encompasses the full multiple scattering behavior that occurs within the solid, while a simple Gaussian beam is used to model the transducer beam pattern. A simplified model restricted to the single-scattering regime is then compared with experimental results for materials of common interest. These results are anticipated to impact ultrasonic nondestructive evaluation and characterization of heterogeneous media.

4:45

4pPA12. Multiple scattering by a random configuration of cylinders in a poro-elastic medium: Properties of the coherent waves. Francine Luppé, Jean-Marc Conoir, and Sébastien Robert (LAUE, UMR CNRS 6068, Université du Havre, place R. Schuman, 76610 Le Havre, France)

An extension of Twersky's formalism is developed in order to obtain the dispersion equation of the coherent waves in a poro-elastic medium in which cylindrical scatterers are randomly placed. The high frequency regime only, where no dispersion effects occur in the absence of scatterers, is investigated. The scatterers lay within a slab-like region of the host medium, and an incident harmonic plane wave gives rise to coherent plane waves in the slab. At normal incidence, the coherent waves are damped, while they are inhomogeneous at other incidence angles. The dispersion equation obtained shows that the longitudinal (fast and slow) coherent waves are coupled, while the shear one obeys the Waterman and Truell's well-known dispersion equation.

5:00

4pPA13. Fresnel approximations for acoustic fields of rectangularly symmetric sources. T. Douglas Mast (Dept. of Biomed. Eng., Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267-0586)

A general approach is presented for determining the acoustic fields of rectangularly symmetric, baffled, time-harmonic sources under the Fresnel approximation. This approach is applicable to a variety of separable source configurations, including uniform, exponential, Gaussian, sinusoidal, and error function surface velocity distributions, with and without focusing in either surface dimension. In each case, the radiated field is given by a formula similar to that for a uniform rectangular source, except for additional scaling of wave number and azimuthal distance parameters. The expressions presented are generalized to three different Fresnel approximations that correspond respectively to diffracted plane waves, diffracted spherical waves, or diffracted cylindrical waves. Numerical results, for several source geometries relevant to ultrasonic applications, show that these expressions accurately depict the radiated pressure fields, except for points very near the radiating aperture. Highest accuracy near the source is obtained by choice of the Fresnel approximation most suited to the source geometry, while the highest accuracy far from the source is obtained by the approximation corresponding to diffracted spherical waves. The methods are suitable for volumetric computations of acoustic fields including focusing, apodization, and attenuation effects.

THURSDAY AFTERNOON, 7 JUNE 2007

GRAND BALLROOM B, 1:30 TO 4:30 P.M.

Session 4pPP

Psychological and Physiological Acoustics: Environmental Sound Research, Spectral Integration, Cochlear Implants, and More (Poster Session)

Marjorie R. Leek, Chair
Portland VA Medical Center, 3710 SW U.S. Veterans Hospital, Portland, OR 97207

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-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

4pPP1. Naturalistic auditory scene analysis in children and adults. Robert Leech, Fred Dick, Jennifer Aydelott (School of Psych., Birkbeck, Univ. of London, London, UK, WC1E 7HX), and Brian Gygi (East Bay Inst. for Res. and Education, Martinez, CA)

In order to make sense of natural auditory environments, the developing child must learn to "navigate" through complex auditory scenes in order to segment out relevant auditory information from irrelevant sounds. This study investigated some of the informational and attentional factors that constrain environmental sound detection in auditory scenes, and how these factors change over development. Thirty-two children (aged 9–12 years) and 16 adults were asked to detect short target environmental sounds (e.g., a dog barking) within longer environmental background

sounds (e.g., a barn) presented dichotically. The target environmental sounds were either congruent (i.e., normally associated with the background) or incongruent (i.e., not normally associated with the background). Subjects heard either a single background presented to both the ipsilateral and contralateral ears or else different background sounds presented to the different ears. Results indicate that children's, but not adults' target detection is substantially less accurate when listening to two auditory scenes than to a single scene presented in both ears, suggesting a developmental shift in listeners' abilities to process multiple information streams. Interestingly, both children and adults find

incongruent sounds more salient than congruent sounds. Furthermore, this incongruence "pop-out" effect is greatly reduced in the presence of the dual background (informational masking) condition.

4pPP2. Noise quality of traffic noise. Klaus Genuit, Sandro Guidati, Sebastian Rossberg, and Andre Fiebig (HEAD acoustics GmbH, Ebertstrasse 30a, 52134 Herzogenrath/Germany, klaus.genuit@head-acoustics.de)

Traffic noise is considered as one major source of noise annoyance in urban context indicated by the increasing number of complaints. The neglect of quality parameters regarding pass-by noise appears as a basic cause for the unreduced annoyance. Several parameters besides the SPL have an important influence on noise evaluation, such as low frequency contributions, temporal aspects or noise patterns. The European research project Quiet City Transport (6FP PL516420) was motivated by the EU Noise Directive 2002/49/EC and is dealing with vehicle pass-by noise and its evaluation in order to identify promising noise mitigation measures. In this framework a quantitative description of noise annoyance using psychoacoustic descriptors was achieved. By means of the detected metric certain perceptual-relevant aspects, such as the difference between diesel and gas, can be reconstructed. Furthermore, the development of a synthesis tool makes the simulation and binaural auralization of pass-by noise, traffic noise as well as specific vehicle noise sources possible. Based on the presented technique the detailed investigation of causes for noise annoyance is carried out. The results will be presented.

4pPP3. Sound source identification by compressive sensing. Raman Arora (Dept. Elec. and Computer Eng., Univ. of Wisconsin, Madison, WI 53705, ramanarora@wisc.edu) and Robert A. Lutfi (Univ. of Wisconsin, Madison, WI 53705, ralutfi@wisc.edu)

Compressive sensing (CS) is an emergent technology that has found increasing application in the areas of broadband signal monitoring and image reconstruction [D. Donoho, IEEE Trans. Info. Theory **52**, 1289–1306 (2006)]. It involves a novel approach to sampling in which the salient information in signals is recovered from the projection of observations onto a small set of randomly-selected basis functions. The major attraction of CS is its capacity for accurate reconstruction based on sparse sampling and little or no prior knowledge of signals. This feature makes it attractive, as well, for application to the problem of sound source identification; an allied task in which sound waveforms as signals are to be classified according to their generating source. In the present paper CS is applied to examples of sound source identification tasks taken from the human psychoacoustics literature. The examples are used to demonstrate in these cases potential advantages of CS classification over traditional decision algorithms that sample at the Nyquist rate. Parallels to the human data are also noted, entertaining speculation as to the role CS classification might play in theoretical thinking about human sound source identification. [Research supported by NIDCD grant 5R01DC006875-02.]

4pPP4. Cochlear implant-mediated perception of non-linguistic sounds. Yell Inverso, Corinne Bickley (Gallaudet Univ., Dept. of Hearing and Speech Lang. Sci., Washington, DC), and Charles Limb (Johns Hopkins Hospital, Baltimore, MD)

Cochlear implants (CI) have provided sound perception to profoundly deaf individuals, revolutionizing the treatment of sensorineural hearing loss. Thus far, the overwhelming majority of test measures to assess implant candidacy, efficacy, and progress are based on speech perception. Non-linguistic speech (NLS) stimuli, by comparison, have received comparatively little attention, despite the importance of NLS sounds for environmental sound awareness. For the prelingually deafened population, in whom speech perception may never be fully realized, NLS perception is an especially critical area deserving further study. The specific aims of this study were to determine the accuracy with which postlingually deafened CI users perceive NLS, and to determine if there is an association between acoustic characteristics of the NLS and ability to identify these sounds using a CI. We used five main categories of NLS: human vocal non-

linguistic, animal/insect, nature/ambient, mechanical/alerting, and musical to comprise a test battery that was given to 10 postlingually deafened CI users. Results of the pilot study showed overall poor ability of CI users to recognize NLS. Furthermore, an association between speech perception and accuracy of NLS was identified. Our results suggest that NLS is a difficult category of complex sound for CI users to perceive.

4pPP5. Spacial cognitive timbre dimensions of physical modeling sounds using multi-dimensional scaling (MDS) techniques. Rolf Bader (Inst. of Musicology, Univ. of Hamburg, 20354 Hamburg, Germany)

The cognitive spacial dimensionality of timbre in the minds of listeners is investigated. Physical models of strings, rods, plates, and membranes and couplings of these obtaining simple musical instruments are used. Here, for each model, eight sounds have been produced by linear change of only one parameter of the physical model. So, i.e., the change of inharmonicity in a string sound modeled as coupling of the string differential equation with a bending stiffness equation by linearly increasing the bending stiffness is investigated. After the judgment procedure, the mean judged similarities of all subjects are fit into an n-dimensional space. The spacial dimensions then have to be interpreted in terms of the physical changes of the sounds distributed over those dimensions. It could be shown that linear changes in inharmonicity of strings do have both, a linear perception as well as sudden phase changes, a finding relevant to psychoacoustic noise perception or room acoustics. Also, the problem of changing plate thickness, different striking points of membranes, or coupling of snare membranes to their vessels with changing cutoff angle of the vessel is investigated. Furthermore, just noticeable differences with timbre changes as well as judgments of very different sounds are discussed.

4pPP6. Practical kurtosis-based blind recovery of a speech source in real-world noise. Matthew D. Kleffner and Douglas L. Jones (Coordinated Sci. Lab., Univ. of Illinois, 1308 West Main St., Urbana, IL 61801, kleffner@uiuc.edu)

In many multi-sensor, single-source applications noise interferes with cleanly recovering the source. Various approaches have been designed to recover sources in interference, but most of them require prior knowledge or assumptions that limit their applicability to real-world environments. We present a novel and practical frequency-domain method for blindly recovering a single, nonstationary, high-kurtosis source in arbitrary low-kurtosis interference using a narrowband kurtosis objective. This method handles convolutional mixing, does not impose a theoretical limit on the number of interferers, and relies only on the kurtosis properties of the source and interference. A kurtosis-based objective is used to compute instantaneous filter weights in each frequency bin, since linear combinations of speech with low-kurtosis interference tend to have lower kurtosis than speech alone. The set of bin-by-bin scale factors needed for the total spatio-temporal recovery filter are computed by using estimated steering vectors to recover the signal as it would sound at a particular sensor. Speech signal-to-interference ratio gains of 5 to 10 dB are demonstrated at low input SIRs in a variety of real-world situations in both a car and a reverberant room.

4pPP7. Amplification and training effects in resolution and cross-spectral integration of redundant cues. Blas Espinoza-Varas (Commun. Sci. & Disord., OUHSC, Oklahoma City, OK 73190), Mohamed Bingabr (UCO, Oklahoma City, OK 73034), and Philip Loizou (Univ. of Texas, Dallas, TX 75083)

Speech recognition of hearing-aid and cochlear-implant users is constrained partly by inefficient integration of information conveyed by multiple frequency channels: owing to cross-channel interference, multichannel recognition accuracy is often lower than the sum of single-channel

accuracies, especially if one channel has impaired sensitivity. With complex tones consisting of a mid- (M) and a high-frequency (H) component (1000 and 3127 Hz), the resolution and cross-spectral integration (CSI) of simultaneous increments in M frequency (IFM) and H duration (ITH) were studied as a function of H amplification and training; M and H impinged, respectively, on normal- and impaired-sensitivity regions of listeners with sensorineural losses. Without amplification, ITH resolution was negligible, the IFM discrimination threshold did not decrease with ITH, and CSI was nil. With H amplification, resolution improved moderately for ITH but decreased for IFM; that is, increasing H audibility interfered with IFM resolution. As a result, the IFM discrimination threshold decreased with ITH but CSI was inefficient and remained so after extensive training. However, ITH resolution and CSI efficiency improved significantly by training listeners to attend to H and respond only on the bases of ITH. The improvement associated with H amplification increased significantly with high-frequency listening training. [Work supported by PHF.]

4pPP8. A correlation feedback model of audio-visual integration.
Sheena Luu and Willy Wong (IBBME, Rosebrugh Bldg., 164 College St., Rm. 407, Univ. of Toronto, Toronto, ON, M5S 3G9, Canada)

To date, a lot of empirical research has been done by many laboratories to investigate how information from different senses interacts to affect perceptual experience. Given the quantity and variety of observations published, it is necessary to begin to organize and integrate the information into a coherent framework. Here a framework is presented in the form of a systems-level model of auditory-visual interaction in the brain. The temporal correlation feedback model is a computational model to illustrate the change in human experience and perception of auditory and visual information as a result of the interaction of the two senses. The model is able to explain many of the effects of audiovisual integration and the characteristics of audiovisual synchrony detection observed in other studies.

4pPP9. Spectral coherence predicts perceptual resilience of speech to temporal distortion.
Christian E. Stilp, Joshua M. Alexander, and Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin—Madison, 1202 W. Johnson St., Madison, WI 53706)

Speech perception is remarkably resilient to signal perturbation. For example, Saberi and Perrott [Nature, **398**, 760 (1999)] found that sentences in which successive segments as long as 130 ms have been temporally reversed are relatively intelligible. They suggested that ultra-low-frequency modulation envelopes (3–8 Hz) are critical to intelligibility. We tested whether information transmitted, not physical units (e.g., ms, Hz), better explains performance. Listeners heard a large cohort of synthesized, seven-syllable sentences at three different speaking rates (2.5, 5.0, 10 syllables/second) with time-reversed segments of multiple durations (20, 40, 80, 160 ms). Intelligibility decreased as duration of time-reversed segments increased, decreasing more slowly for slower (longer) sentences. When intelligibility is plotted to reflect proportion of message (not time) that was time reversed, data for all three rates collapse onto a single curve. To estimate potential information across time, we used spectral coherence to quantify the degree to which successive portions of the signal are similar to, or can be predicted from, a given spectral slice with no prior expectations such as knowledge of speech or language. Spectral coherence alone accounts substantially for performance. Information transmission, not physical measures of time or frequency, is more useful for understanding perception of speech. [Work supported by NIDCD.]

4pPP10. Modulation interference in speech recognition by cochlear implant users.
Bom Jun Kwon (Dept. of Commun. Sci. and Disord., Univ. of Utah, Salt Lake City, UT 84112) and Peggy B. Nelson (Univ. of Minnesota, Minneapolis, MN 55455)

When listening to speech in a background masker, normal-hearing listeners take advantage of envelope fluctuations or dips in the masker (or the moments that the SNR is instantaneously favorable). Cochlear implant listeners, however, do not exhibit such ability. In a previous study [Nelson *et al.*, J. Acoust. Soc. Am. **113**, 961–968 (2003)], speech recognition performance with modulated maskers was similar or slightly worse than with a steady noise. Clinical observation indicates that implant listeners usually have more difficulty understanding speech in modulated backgrounds. In the present study, the recognition of IEEE sentences by Nucleus recipients was measured in a variety of backgrounds. In tightly controlled conditions via direct stimulation, performance is often substantially poorer with modulated backgrounds (the differences in score are as large as 20–30%), strongly indicating that they are subject to modulation interference [B. J. Kwon and C. W. Turner, J. Acoust. Soc. Am. **110**, 1130–1140 (2001)]. A mixture of speech and modulated backgrounds might be perceived as one distorted signal, rather than two signals, compromising the identity of the speech, as perceptual segregation of signals is very challenging. A future implant system should incorporate a strategy to provide multiple input sounds without sacrificing the sound identity.

4pPP11. Simulating bipolar and monopolar cochlear-implant stimulation in speech recognition by normal-hearing listeners.
Mohamed Bingabr (Dept. of Phys. & Elec. Eng., UCO, Oklahoma City, OK 73034), Blas Espinoza-Varas (OUHSC, Oklahoma City, OK 73190), Philip C. Loizou (Univ. of Texas, Dallas, TX 75083), and Kenneth Hawkins (UCO, Oklahoma City, OK 73034)

An algorithm was developed to simulate cochlear-implant (CI) bipolar and monopolar stimulation (BS and MS) in speech recognition by normal-hearing listeners. In MS, the active electrode is inside and the return is outside the cochlea, and electrical current decays at 0.5 dB/mm. In BS, both electrodes are adjacent inside the cochlea and current decays at 4.0 dB/mm to either side of the active electrode [Bruce *et al.*, IEEE Trans. Biomed. Eng. **46**, 617 (1999)]. Owing to the faster decay rate, spatial (spectral) resolution is higher with BS than with MS. In recognition of CNC sentences, decay rate was simulated as spectral smearing. Speech was processed through 6, 11, or 22 pass bands, rectified, and low-pass filtered (200 Hz). The passbands amplitude envelopes modulated the amplitude of noise bands equal in center frequency to the passbands. The spectral envelope of each noise band decayed to either side of the center frequency at a rate matching the MS or BS current decay (i.e., noise-band bandwidth was proportional to decay rate). With the summed noise bands, recognition was studied as a function of number of pass bands. [Work supported by NIH.]

4pPP12. Speech intelligibility in noise with single-microphone noise reduction implemented in 9-, 16-, and 24-channel compressors.
Michael J. Nilsson, Michelle L. Hicks, Robert M. Ghent, Jr., and Victor H. Bray, Jr (Sonic Innovations, Auditory Res. Lab, 2795 East Cottonwood Pkwy., Ste. 660, Salt Lake City, UT 84121)

Reception Threshold for Sentences as measured by a modified version of the HINT test [M. J. Nilsson *et al.*, J. Acoust. Soc. Am. **95**(2), 1085–1099 (1984)] were collected on hearing-impaired listeners fit binaurally with digital hearing aids incorporating a 9-channel, 16-channel, and 24-channel spectral subtraction-like technique of single-microphone noise reduction. Thresholds were measured in noise presented in a two-dimensionally diffuse soundfield with the subjects listening unaided, aided without noise reduction, and aided with noise reduction. The ability of noise reduction to reduce the level of steady state noise was quantified using 2 cm3 coupler measures with flat linear fittings (to measure the maximum attenuation possible) as well as using each of the listeners' prescribed fittings (to better relate to the changes in performance measured

in the sound field). Previously measured benefit from noise reduction [(M. J. Nilsson *et al.*, J. Acoust. Soc. Am. (2000)] is hypothesized to occur from frequency-specific gain manipulations that increases gain in frequencies where speech is the dominant signal and decreases gain in frequencies where noise is the dominant signal. Additional benefits from smaller, independent channel structures are therefore expected from the 16 and 24 channel systems.

4pPP13. Validation of the noise reduction index (NRI) as an estimate of the signal-to-noise ratio change of a mixed signal by a hearing aid.
Robert M. Ghent, Michael J. Nilsson, and Victor H. Bray (Sonic Innovations, Inc., Auditory Res. Dept., 2795 E. Cottonwood Pkwy., Ste. 660, Salt Lake City, UT 84121)

The NRI is an extension of an idea discussed by Licklider [J. Acoust. Soc. Am. **20**, 150–159 (1948)] and implemented in practice by Hagerman and Olofssen [Acta Acoustica **90**, 356–361 (2004)] to investigate the conventional wisdom that asserts that a hearing aid cannot fundamentally alter (improve) the input SNR using a single-microphone signal processing algorithm such as digital noise reduction. The present research explores the validity of the NRI by using it to estimate the change in input SNR through a variety of audio devices, including linear hearing aids (with and without directional microphones), a directional microphone designed for noise rejection on a concert stage, an ear trumpet, and a multi-channel hearing aid with a digital noise reduction algorithm both engaged and disengaged. Results indicate that the change in SNR through an audio device can be negative as well as positive, that the NRI is a robust and valid method for estimating this change, and that single-microphone noise reduction algorithms can improve the input SNR when properly designed. The test setup and methodology, along with some pitfalls to avoid when performing these measurements, will be discussed.

4pPP14. Speech perception from a crudely quantized spectrogram: A figure-ground analogy.
Douglas Brungart, Nandini Iyer, and Brian Simpson (Air Force Res. Lab., 2610 Seventh St., WPAFB, OH 45433, douglas.brungart@wpafb.af.mil)

Recent research into auditory chimeras has shown that comprehensible speech can be generated by superimposing the frequency-dependent modulation envelope of a target speech signal onto the fine structure of music, speech, or almost any other broadband waveform. In this experiment, we show that, in certain cases, intelligibility can be achieved with an even simpler stimulus that preserves only the rough outline of the energy distribution of the target speech waveform. This stimulus was generated by 1) dividing the target speech spectrogram into discrete tiles that were 1/3rd octave band wide and 7.8 ms long; 2) identifying the minimum number of tiles necessary to capture 90% of the energy in the target waveform; and 3) filtering a uniform, broadband noise to activate only those tiles that met the threshold energy criterion in the original stimulus. The resulting stimulus captures the overall outline of the distribution of energy in the target speech, but contains no information about the relative distribution of energy within the spectrogram. In this paper, we discuss the use of this stimulus in a series of auditory figure-ground experiments where the target speech signal differs from the background by only a single parameter (relative level, ITD, ILD, etc.).

4pPP15. Transformation of live versus recorded speech from the mouth to the open or occluded ear.
Dragana Barac-Cikoja, Linda Kozma-Spytek, and Stephanie Adamovich (Gallaudet Univ., 800 Florida Ave. NE, Washington, DC 20002, dragana.barac-cikoja@gallaudet.edu)

Auditory self-monitoring studies require the participants' speech feedback to be manipulated experimentally and then presented to the ear. The acoustic levels of the altered speech must be the same as those under normal speech feedback conditions when speech is transmitted to the ear live, directly from the mouth. Therefore, it is critical to understand the transfer function between the mouth and the eardrum. To further this understanding, live and played-back self-generated speech for two female and two male individuals were recorded using a probe tube microphone inserted into the ear canal at a fixed distance from the tragus. The played-back speech samples were delivered via a B&K artificial mouth, at the sound pressure level calibrated to match the level at the lips. The artificial mouth was placed offset to the participants right side and in line with the *x* and *z* axes of his/her mouth position. Recordings were conducted with the ears either open or sealed with ER-3A insert earphones. In addition, the level and content of speech were systematically varied. The equivalent sound level and long-term average spectra of the corresponding pairs of speech samples were analyzed as a function of the experimental manipulations.

4pPP16. An active noise cancellation system to reduce the intense noise caused by magnetic resonance imaging.
Deborah Hall, John Chambers (MRC Inst. of Hearing Res., University Park, Nottingham, NG7 2RD, UK), Kimberly Quinn (Nottingham Trent Univ., Nottingham, NG11 8NS, UK), and Michael Akeroyd (MRC Inst. of Hearing Res., Glasgow, G31 2ER, UK)

Magnetic resonance imaging (MRI) has become the dominant technique for noninvasive measurements of central auditory function in humans, but its operation generates intense acoustic noise (some 115 decibels). Although this can typically be attenuated by about 30 decibels with ear defenders, there is considerable scope for additional benefit. We took advantage of two features of the sound: (i) its predictable temporal onset, and (ii) its fixed power spectrum (which is dominated by intense harmonic components at low frequencies) to engineer an active noise cancellation (ANC) system to achieve further noise reduction. We measured its perceptual effectiveness by collecting detection thresholds for a tonal signal, masked by the image acquisition noise of the scanner (Philips 3 Tesla Intera) with and without the ANC operating. We presented signal frequencies centered at the peaks of the objective cancellation (150, 300, and 600 Hz), plus 450 Hz (a control frequency achieving no significant acoustic cancellation). Across four listeners, the ANC system gave average benefits of 4, 11, 15 dB, respectively, at the peaks and 10 dB at 450 Hz. These results confirm that the system offers substantial benefits for audibility in MR research. Moreover, the ANC equipment does not interfere with image quality.

4pPP17. Effect of level variation on intensity resolution in intensity-discrimination and increment-detection paradigms.
Walt Jesteadt and Donna L. Neff (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

In studies of profile analysis, the level of the pedestal and increment have been varied together from interval to interval by up to 40 dB to reduce the likelihood that subjects could perform the task by comparing the overall level in the two observation intervals. This range of rove was chosen based on a calculation by Green [*Profile Analysis* (Oxford, New York, 1988)]. In the present study, adaptive thresholds for intensity resolution were obtained in intensity discrimination (pedestal and increment 100 ms) and increment detection (pedestal 100 ms, increment 300 ms) paradigms as a function of the range of interval-to-interval level variation (0 to 50 dB in 10-dB steps). In roving conditions, the levels of pedestal and increment varied in parallel. In jittered conditions, only pedestal level varied from interval to interval. Increments were 2-kHz sinusoids. Pedestals were 2-kHz sinusoids or narrow-band noises centered on 2 kHz. Effects of level variation were greater in roving than in jittered conditions. Increment detection was unaffected by roving level, whereas thresholds for intensity discrimination were poorer than Green's best-case prediction by a relatively constant amount for all rove ranges. [Work supported by NIH-NIDCD.]

Session 4pSC

Speech Communication: Vowel Production and Perception (Poster Session)

Stephen M. Tasko, Chair

Western Michigan Univ., Speech Pathology and Audiology, 1903 W Michigan Ave., Kalamazoo, MI 49008-5355

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-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

4pSC1. Spatial-temporal patterns of vowel articulation: A cross-speaker study. Stephen Tasko (Dept. of Speech Pathol. and Audiol., Western Michigan Univ., 1903 W. Michigan Ave., Kalamazoo, MI 49008-5355, stephen.tasko@wmich.edu)

An important goal in speech production research is to establish a quantitative description of normal or typical articulatory behavior. This description should be expressed in a compact but meaningful way and be robust enough to generalize across a large number of speakers who differ in physical size, gender, age, etc. This is a significant challenge given the numerous degrees of freedom within the articulatory system and the range of speaker variability. However, Hashi *et al.* [J. Acoust. Soc. Am. **104**, 2426–2437 (1998)] used x-ray microbeam data to develop a procedure for normalizing vocal tract size and shape. This procedure reduced cross-speaker variation in the articulatory postures of vowels. The authors limited their analysis to a single time slice at the acoustic steady state. There is little available information regarding speaker-general time-varying articulatory patterns associated with vowel production. This study will address this issue. Fifty-three healthy speakers from the University of Wisconsin x-ray microbeam database served as subjects. The speech task consisted of productions of American English vowels in a /sVd/ context. Analysis focused on the movement of a number of vocal tract markers over the duration of each vowel. Vowel-specific spatial temporal patterns common to the speaker group will be described.

4pSC2. Predicting pharyngeal diameter from anterior tongue pellet positions in Swedish. Michel T-T. Jackson and Richard S. McGowan (CReSS LLC, 1 Seaborn Pl., Lexington, MA 02420)

One goal of phonetics is to quantitatively understand articulation in human speech production. As part of this goal, we are interested in the extent to which some vocal tract dimensions, specifically midsagittal pharyngeal diameters, are predictable from flesh points of the anterior tongue. In a reanalysis of x rays of speakers producing Swedish vowels, we modified previous procedures [Whalen *et al.*, J. Speech, Lang. Hear. Res. **42**, 592–603 (1999); Nix *et al.*, J. Acoust. Soc. Am. **99**, 3707–3717 (1996)] to predict pharyngeal diameters from pseudo-pellet positions. We used principal component analysis to reduce the number of pellet degrees of freedom from eight to three prior to applying linear regression from these three independent variables to each of several dependent pharyngeal diameters. Except for the regions around the laryngopharynx and uvula, the pharynx dimensions predictable from linear regressions were significant at the $p < 0.05$ level. Numerical experiments show that it is crucial to reduce the number of independent variables in tests of statistical significance. [Work supported by Grant No. NIDCD-001247 to CReSS LLC.]

4pSC3. Vowel identification and vowel space characteristics. Amy Neel (Dept. of Speech & Hearing Sci., MSC01 1195, 1 Univ. of New Mexico, Albuquerque, NM 87131) and Jean Andruski (Wayne State Univ., Detroit, MI 48202)

Acoustic characteristics of ten vowels produced by 45 men and 48 women from the Hillenbrand *et al.* (1995) study were correlated with identification accuracy. Global (mean f0, F1 and F2, duration, and amount of formant movement) and distinctive measures (vowel space area, mean distance among vowels, f0, F1 and F2 ranges, duration ratio between long and short vowels, and dynamic ratio between dynamic and static vowels) were used to predict identification scores. Global and distinctive measures accounted for less than one-fourth of variance in identification scores: vowel space area alone accounted for 9% to 12% of variance. Differences in vowel identification across talkers were largely due to poor identification of two spectrally similar vowel pairs /ae/-/eh/ and /uh/-/ah/. Results of acoustic analysis and goodness ratings for well-identified and poorly identified versions of these vowels will be presented. Preliminary analysis revealed that well-identified vowels differed from poorly identified tokens not only in static formant frequencies but also in duration and amount of formant movement over time. A talker's ability to create distinctions among neighboring vowels is more important in determining vowel intelligibility than overall measures such as vowel space area.

4pSC4. Exploring the natural referent vowel hypothesis in infant perception experiments. Ocke-Schwen Bohn (English Dept., Arhus Univ., DK-8000 Arhus C, Denmark)

Directional asymmetries in infant vowel perception studies have led Polka and Bohn [Speech Commun. **41**, 221–231 (2003)] to posit the Natural Referent Vowel hypothesis (NRV) according to which vowels that are relatively more peripheral in the F1/F2 space are perceptually privileged *vis-a-vis* less peripheral vowels. This bias has been observed in preference and in change/no change experiments with infants and with adult non-native listeners. NRV was further tested in headturn (change/no change) experiments with Danish-learning infants. With one important exception, results support the NRV by showing that the predicted asymmetries exist in areas of the vowel space that have not been previously explored. The exception was observed for the Danish front vowel pair /e/-/ø/, where NRV incorrectly predicts that a change from the more peripheral /e/ to the less peripheral /ø/ is harder to discriminate than a change in the opposite direction. Interestingly, the unexpected asymmetry was observed only with infants up to the age of 7.5 months. Older infants showed a nonsignificant trend for an asymmetry in the opposite direction, as predicted by NRV. A modification of NRV to incorporate this finding will be proposed. [Research supported by Grant 25-01-0557 from the Danish Research Council for the Humanities (FKK).]

4p THU. PM

3187   J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America   3187

4pSC5. Identification of brief vowel sequences by young and elderly listeners. Diane Kewley-Port, Larry Humes, and Daniel Fogerty (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405, kewley@indiana.edu)

Aging generally has a negative impact on temporal processing. Four experiments using vowel sequences are presented from a large series of temporal processing studies of age-related changes across three sensory modalities. These four experiments used adaptive tracking to determine the smallest temporal difference between the onsets of vowels (SOA) that preserved identification of either the vowel sequences or ear presented. Vowels in the four words (pit, pet, pot, put) from a male speaker were processed using STRAIGHT to have the same length (70 ms) and F0 (100 Hz) and equated for rms level. Young ($N=40$) and elderly ($N=40$; 60–79 years) adults identified the vowels in isolation with better than 90% accuracy. Preliminary analyses based on 40% of the data are summarized here. Baseline performance by young listeners in the monaural, two-vowel sequence task was $SOA=20$ ms. For the other three conditions (monaural four-vowel sequences, dichotic two-vowel sequences, and dichotic two-vowel sequences for ear presented), SOA values increased by three to four times. This pattern of SOA values was similar for the elderly listeners. However, as expected, SOA values were significantly longer for the elderly group, but only by a small percentage ($<30\%$). [Work supported by NIA R01 AG022334-02.]

4pSC6. The development of relative amplitude changes across acoustic discontinuities in nasal+vowel syllables. Meredith A. Oakey and Ralph N. Ohde (Vanderbilt Univ., Vanderbilt Bill Wilkerson Ctr., 1215 21st Ave. South, Rm. 8310, Nashville, TN 37232-8242)

In adult productions of nasal+vowel syllables, relative amplitude changes occur in various frequency regions across acoustic discontinuities and provide important cues as to the place of articulation of nasal consonants. The purpose of this study was to investigate the development of relative amplitude changes in children as potential acoustic correlates to place of articulation. Four age groups of eight participants each (3, 5, 7, adult) served as speakers. Participants produced five productions each of CV syllables comprised of [m] and [n] in the context of four vowels ([i æ u ɑ]). These syllables were segmented into approximately 25 ms segments of the murmur and 25 ms segments of the vowel bordering the period of discontinuity. The relative amplitude changes in low- and high-frequency ranges from the murmur to the vowel were determined using summed amplitude information from fast Fourier transform (FFT) analyses. Previous research showed systematic differences in relative amplitude between [m] and [n] in adult syllable productions, but only marginal differences were observed in very young children's speech. The relative amplitude property, a landmark cue representing place of articulation of nasal consonants, is predicted to develop with age based on these preliminary analyses. [Work supported by NIH and an ASHA SPARC Award.]

4pSC7. An acoustical study of English constant-vowel (CV) syllables. Byunggon Yang (English Education Dept., Pusan Natl. Univ., 30 Changjundong Keumjunggu, Pusan 609-735, South Korea, bgyang@pusan.ac.kr)

This study examined acoustic characteristics of 900 CV syllables produced by five English native speakers downloaded from http://www.nsi.edu/users/patel/download.html. Those target syllables were produced quite synchronously between the syllable /ba/ twenty times. The syllables were segmented and normalized by the maximum intensity value of each syllable and were divided into consonant or vowel sections by a few visible acoustic criteria. Intensity values were collected at 100 relative time points per syllable. Also, cumulative intensity values and consonant and vowel durations along with the ratio of the onset consonant to each syllable were measured using Praat scripts. Results showed as follows: First, the consonantal section amounted to a quarter of the syllable in terms of both the cumulative intensity and duration. Second, the consonantal ratio by the cumulative intensity was similar to that by the duration.

Finally, the sum of the cumulative intensity values in each syllable partially coincided with the order of consonant groups on the current sonority scale while the vowel section of each syllable showed almost comparable intensity curve. Further studies would be desirable on more reliable acoustical measurements and sophisticated perceptual experiments on the English syllables.

4pSC8. Modeling vowel normalization and sound perception as sequential processes. Ricardo A. Hoffmann Bion and Paola Escudero (Inst. of Phonetic Sci., Univ. of Amsterdam, The Netherlands, escudero@uva.nl)

This study constitutes the first attempt at combining vowel normalization procedures with the linguistic perception framework of stochastic Optimality Theory and the Gradual Learning Algorithm (Boersma, 1998). Towards this end, virtual learners possessing different normalization procedures, and a control learner with no normalization procedure, were trained to perceive Brazilian Portuguese and American English vowels. The only parameters fed into the model were the first two formants of the vowels. The simulations assume that: (i) vowel normalization occurs prior to vowel categorization, (ii) each learner has acquired a different normalization algorithm at the time of the simulations, and (iii) perceptual learning takes place when mismatches between the word intended by the speaker and the word perceived by the virtual listener occur (Escudero & Boersma 2004). Our results show that learners equipped with normalization algorithms outperformed the control learner, obtaining accuracy scores from 20 to 30% higher. When equipped with vowel-extrinsic procedures, learners managed to reach more than 90% of correct responses. Thus, a formal model in which normalization and sound perception are implemented as two sequential processes delivers the expected results, as it accurately models vowel perception even when the training and testing sets have speakers with different ages and gender.

4pSC9. The value of F0, F3, and F4 in identifying disguised voices. Jean E. Andruski, Nikki Brugnone, and Aaron Meyers (Dept. of Commun. Sci. & Disord., 207 Rackham Bldg., 60 Farnsworth Ave., Detroit, MI 48202, ag0611@wayne.edu)

This study examines the value of F0, F3, and F4 for identifying speakers from a group of ten male and female speakers when the speakers deliberately changed their F0 and/or their articulatory patterns. The speakers were recorded producing a short passage in their normal speaking voice, a lower than normal speaking voice, and using a vocal disguise of their choice. Although F1 and F2 are strongly affected by articulation, the higher formants show relatively little effect of articulation. This relative stability may make them more useful than F1 and F2 for examining speaker identity. While formants reflect vocal tract cavity dimensions, F0 reflects a different aspect of speaker identity, namely the mass and stiffness of the speaker's vocal folds. Like F3 and F4, F0 shows relatively little effect from articulation. However, speakers can voluntarily shift F0 into a range that is different from what they normally use. Measurements of F0, F3, and F4 were taken on selected stressed vowels from each recording condition. Measurements from the normal voice condition were used to train a discriminant analysis, which then classified vowels from the low F0 and vocal disguise conditions by speaker. Accuracy of the classification results will be presented and discussed.

4pSC10. Tongue action in silence: Japanese voiceless vowels. Diana Archangeli, Adam Baker, Alina Twist, and Yuko Watanabe (Univ. of Arizona, Douglass Bldg., Rm. 200E, Tucson, AZ 85721, yukow@email.arizona.edu)

The study presents acoustic and articulatory analyses of the Japanese voiceless vowels [(i)] and [(u)] arguing Japanese listeners have internal representations for the vowels and perceive them from articulatory information while physical properties of vowels are absent. Ogasawara (2006)

reports that there is no formant band in Japanese voiceless vowels. It is an open question as to whether there is an articulatory gesture associated with the voiceless vowels, and whether this gesture is the same or different as that of voiced vowels. To answer the question, we investigated production of voiceless vowels by a native Japanese speaker. We created 11 pairs of nonsense words such as [hogito] vs [hok(i)to] and [hoguto] vs [hok(u)to]. We recorded the speakers production of the stimuli with the voiceless vowels and their voiced counterparts using the ultrasound device to capture tongue images. The same recording was examined for acoustic analysis. An investigation of the ultrasound images reveals there are tongue gestures for the voiceless vowels, movements differ depending on which voiceless vowel it is, and the movements parallel those of the voiced counterparts. These results argue in favor of the view that Japanese speakers internal representations of words include a representation of distinct voiceless vowels.

4pSC11. Are vowels in African American English affected by Southern vowel shift? Yolanda F. Holt, Ewa Jacewicz, and Robert A. Fox (Speech Percept. and Acoust. Labs., Speech and Hearing Sci., Ohio State, 1070 Carmack Rd., Columbus, OH 43210-1002, holt.174@osu.edu)

There is a debate regarding whether the variety of speech called African American Vernacular English (AAVE) participates in any of the vowel shifts identified in today's American English. Because pronunciation of vowels in AAVE has never been studied with sufficient experimental rigor, it is unknown whether the vowel system of AAVE differs from those of closely-related regional variants of American English. The present experiment examines the apparent resistance of AAVE to assimilation into Southern American English (SAE) and seeks to determine the presence or absence of Southern Vowel Shift in the speech of speakers of AAVE. Acoustic measures of AAVE vowels produced by male speakers born and raised in Statesville, NC (Piedmont region) are compared with the same measures of SAE vowels spoken by European-American male speakers living in western North Carolina. Acoustic measures (obtained from fourteen monophthongs and diphthongs in /h_d/ context) included dynamic formant pattern, extended vowel space area for each speaker, the amount of frequency change for F1 and F2 at two temporal points located close to vowel onset and offset (vector length), and vowel duration. These measures will establish both systemic and vowel inherent characteristics within these two varieties of English. Dialectal differences will be discussed. [Work supported by NIH NIDCD R01 DC006871.]

4pSC12. Vocal tract normalization for vowel recognition. Robert A. Houde (Ctr. for Commun. Res., 125 Tech Park Dr., Rochester, NY 14623, rahoude@gmail.com) and James M. Hillenbrand (Western Michigan Univ., Kalamazoo, MI 49008)

The vocal tract normalization problem arises from the fact that the spectrum envelopes of like vowels spoken by talkers with different vocal tract lengths appear as the same spectral pattern shifted along a log frequency scale. A narrow-band model of vowel perception was developed in which vowels are identified by comparing input narrow-band vowel spectra with broad-band pitch-independent vowel templates [J. M. Hillenbrand and R. A. Houde, J. Acoust. Soc. Am. 113, 1044–1055 (2003)]. An earlier evaluation of this model was carried out on a large database of vowels spoken by men, women, and children. In this evaluation the vocal tract normalization problem was bypassed by constructing separate templates for men, women, and children. The present study extended this work by examining two approaches to the vocal tract normalization problem: (1) An explicit normalization approach in which the spectral pattern of each input token was normalized by a factor based on the pitch of the token and (2) an approach in which templates were assembled by summing together spectral patterns that were sufficiently similar, thus creating multiple tem-

plates for each vowel. Performance of the model with these two approaches to vocal tract normalization was compared to listening tests on a large database of vowels spoken by men, women, and children.

4pSC13. Salience of dynamic virtual formants in diphthongs. Robert Allen Fox, Ewa Jacewicz, and Chiung-Yun Chang (Speech Percept. and Acoust. Labs., Speech and Hearing Sci., Ohio State, 1070 Carmack Rd., Columbus, OH 43210-1002, fox.2@osu.edu)

Variations in the quality of a synthesized vowel can be produced by changing the frequency of its steady-state formants or the amplitude of closely-spaced static spectral components—the latter effect a function of auditory spectral integration. Recent studies have demonstrated that dynamic modification of the amplitude ratios of these spectral components can give rise to the perception of virtual diphthongs. This study examines the extent to which spectral integration of these dynamic components is uniform across a range of frequency separations of these spectral components (previous work [Fox et al., "Title," J. Acoust. Soc. Am. 120, 3252 (2006)] demonstrated the integration of static components beyond 3.5 bark). Two synthetic vowel series were created by modifying the rise or fall of F2 producing [əi]-[əː]-[əu] and [iə]-[əː]-[uə] series. Next, several different series were created in which this dynamic F2 was replaced by two static spectral components whose amplitude ratios were varied over time (virtual F2 glides). The spectral separation of these two components varied from 2.0 to 4.0 bark. A single-interval 2AFC identification task was used to obtain estimates of vowel quality. Results will be discussed in terms of whether the spectral integration effects decline as the frequency separation between components increases. [Work supported by NIDCD R01DC00679-01A1.]

4pSC14. Formant amplitude in the perception of /i/ and /u/. Michael Kiefte, Teresa Enright, and Lacey Marshall (School of Human Commun. Disord., Dalhousie Univ., Halifax, NS B3H 1R2, Canada)

Although recent evidence reconfirmed the importance of spectral peaks in vowel identification [M. Kiefte and K.R. Kluender, J. Acoust. Soc. Am. 117, 1395–1404 (2005)], the roles of formant bandwidth and amplitude in perception have not been firmly established. Although several studies have shown changes in vowel categorization with changes in bandwidth and amplitude, such manipulations often result in peaks that have very little local spectral contrast or which may be perceptually masked by other spectral prominences. This study compares local contrast and masking of spectral peaks in vowel perception. A stimulus continuum ranging from /i/ to /u/ in both the amplitude of the second formant as well as that of higher formants was presented to listeners for identification. Both full-spectrum and incomplete-spectrum stimuli were used to demonstrate the effects of local spectral contrast and simultaneous masking. A second experiment was used to determine if listeners were able to detect the presence of the second spectral peak in the same stimuli. Results from both experiments indicated that both masking and spectral contrast play a role in vowel perception.

4pSC15. Effects of dialect variation on speeded word classification. Cynthia G. Clopper (Dept. of Linguist., Ohio State Univ., 1712 Neil Ave., Columbus, OH 43210, clopper.1@osu.edu)

Dialect variation has been found to significantly affect speech intelligibility in noise. The current experiment was designed to examine the effects of dialect variation on the processing of a single phonemic contrast. Participants were presented with tokens of the words "bad" and "bed" produced by talkers from Chicago and Indianapolis. The vowels in the Chicago tokens were shifted according to the Northern Cities shift (raised and fronted "bad," lowered and backed "bed"). The Indianapolis tokens were unshifted. The participants were asked to classify each token as "bad" or "bed" as quickly as possible in three conditions: Chicago tokens only, Indianapolis tokens only, or mixed Chicago and Indianapolis tokens.

While overall accuracy was high in all three conditions, participants were slower in the Chicago-only and mixed conditions than in the Indianapolis-only condition. In addition, participants responded more slowly to the Indianapolis tokens in the mixed condition than in the Indianapolis-only condition, but response times to the Chicago tokens did not differ across the two conditions. These results suggest that acoustic-phonetic dialect differences interfere with lexical processing and that global effects of dialect variation on speech intelligibility may reflect confusions between specific pairs of phonemes as a result of subphonemic vowel shifts.

4pSC16. Cross-linguistics differences in non-speech perception.
Molly Babel and Keith Johnson (Univ. of California, Berkeley, Berkeley, CA 94709-2650, mbabel@berkeley.edu)

This research explores listeners perception of non-speech stimuli across languages. We will report on two experiments conducted with the same stimuli. The stimuli were synthesized from naturally produced vowel-fricative-vowel sequences consisting the fricatives [f, th, s, sh, x, h], and the vowel environments [i_i, a_a, u_u]. The vowels were synthesized using sine wave synthesis and appended onto the natural fricative. Tokens will be presented to two groups of Dutch and English listeners over headphones with a 100 ms interval. Listeners will be told that the stimuli is non-speech. The first experiment is a simple AX discrimination task where the listener decides whether the tokens were same or different. The second experiment is a rating task where the listener is asked to rate the similarity of the two tokens on a five-point equal interval scale. It is predicted that results of both experiments will be the same for both groups of listeners. These results will be compared to the results of the same experiment with non-synthesized speech [K. Johnson and M. Babel, J. Acoust. Soc. Am. **120**, 3292 (2006)]. [Work supported by the NIH.]

THURSDAY AFTERNOON, 7 JUNE 2007

GRAND BALLROOM C, 1:30 TO 4:35 P.M.

Session 4pSP

Signal Processing in Acoustics, Acoustical Oceanography, and Underwater Acoustics: Signal Processing and Uncertainty II

Lee Culver, Cochair
Pennsylvania State Univ., Applied Research Lab., P. O. Box 30, State College, PA 16804-0030

Ning Xiang, Cochair
Rensselaer Polytechnic Inst., Architecture, 110 8th St., Greene Bldg., Troy, NY 12180

Invited Papers

1:30

4pSP1. Some speculations on source localization in uncertain ocean environments. Steven Finette (Acoust. Div. Naval Res. Lab., Washington, DC 20375)

Incomplete environmental knowledge is a factor that can adversely affect the prediction capability of physics-based signal processors to localize acoustic sources in ocean waveguides. Environmental uncertainty effectively introduces unwanted additional degrees of freedom into the signal processing scheme. Including a quantitative measure of this incomplete information into the processor can be accomplished by considering the influence of alternative environments, each weighted by their probability of occurrence. An approach is outlined to account for environmental uncertainty in matched field processing by the use of polynomial chaos basis expansions to represent the replica vectors in terms of stochastic Green's functions. These bases are interpreted as representing the effect of introducing additional degrees of freedom due to environmental uncertainty, because the pressure field replicas are specified as random processes. In this manner, a processor correlates the measured pressure field with a set of statistically weighted propagators between a replica source location and a sensor array. [Work sponsored by ONR through NRL base funding.]

1:50

4pSP2. Wigner approximation for filtered random functions and wave propagation. Patrick J. Loughlin (Dept. of Elec. and Computer Eng., Univ. of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15261, loughlin@engr.pitt.edu)

We consider the time-varying (Wigner) spectrum of a filtered signal and derive an approximation that shows the effects of the magnitude and phase of the filter on the time-varying spectrum of the original signal. Both the deterministic and random cases are considered. Application to wave propagation with dispersion and damping is discussed.

4pSP3. Ensemble-based Bayesian detection and tracking. Brian La Cour (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

This paper introduces a novel approach to Bayesian tracking in which the posterior distribution is represented by a random sample or ensemble of possible target states. Such a representation may be viewed as a special case of the traditional particle filter approach wherein all particles maintain strictly uniform weights. Measurement updates are performed using a Markov chain Monte Carlo technique, which has been adapted to use multiple chains and a variable pseudo temperature akin to that used in simulated annealing. The general formalism is illustrated in an example of sonar-based target tracking for antisubmarine warfare. For this example, specific motion models and likelihood functions are developed for both target and clutter hypotheses. The technique is examined in the context of results from a recent multistatic seatrial using an echo repeater target and compared against those of a traditional particle filter. [This work was supported by ONR Contract No. N00014-00-G-0450-21.]

4pSP4. Use of polynomial chaos expansions in the design of uncertainty-robust matched field processing algorithms. Kevin D. LePage (Naval Res. Lab., Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375)

The polynomial chaos (PC) expansion method has been recently applied to estimating the statistical properties of underwater acoustic propagation in the presence of environmental uncertainty. Here we use PC estimates of the field covariance structure to design uncertainty robust match field processing (MFP) weights using Krolik's minimum variance beamformer with sound-speed perturbation constraints (MV-SPC) [J. Acoust. Soc. Am. **92** (3), 1408–1419 (1992)]. The idea behind the MV-SPC beamformer is to realize much of the high sidelobe rejection of MV processors in environments with environmental variability by opening up the signal model to include uncertainty effects. Here we compare the performance of MV-SPC designed with an adiabatic signal uncertainty model to the same beamformer designed with the PC signal model for realizations of the signal vector obtained with fully coupled propagation models, with results showing the superiority of the PC approach. [Work supported by ONR.]

Contributed Papers

4pSP5. A ray-based blind channel estimation method in ocean sound channels. Application to high-frequency underwater telecommunications. Karim G. Sabra (Marine Physical Lab., Scripps Inst. of Oceanogr., UC San Diego, San Diego, CA 92093-0238, ksabra@mpl.ucsd.edu), David R. Dowling (Univ. of Michigan, Ann Arbor, MI), H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal (UC San Diego, San Diego, CA 92093-0238), J. M. Stevenson, and Piero Guerrini (NATO Undersea Res. Ctr., La Spezia, Italy)

A signal that travels through a sound channel is typically distorted when recorded by a remote listener because of interference arising from multiple propagation paths. We developed a ray-based blind channel estimation technique in ocean sound channels, i.e., using only records of the distorted signal along a transducer array. The technique yields broadband estimates of the source-to-array impulse response (or Green's functions) synthesized directly from the recorded distorted signals. The technique relies on separating out the contribution of a propagating ray path having a phase that depends linearly on frequency. This technique was used to decode high-frequency underwater telecommunications sequences collected in 2005 off the West Coast of Italy, without any *a priori* knowledge of the channel impulse response. [Work supported by ONR.]

4pSP6. Probabilistic descriptions of low-frequency passive sonar signals. Colin W. Jemmott, Jeffrey A. Ballard, R. Lee Culver, and Leon H. Sibul (Appl. Res. Lab. and Grad. Program in Acoust., The Penn State Univ., P.O. Box 30, State College, PA 16804)

The overall goal of our research is to develop a method whereby knowledge of environmental variability is incorporated into a passive sonar detector. The Estimated Ocean Detector (EOD) is a maximum likelihood receiver that incorporates statistical knowledge of signal uncertainty in the form of signal parameter probability density functions (pdf's). The structure and performance of the EOD depends fundamentally on accurate probabilistic models of received signal and noise parameters, the so-called prior distributions. Previous work has concentrated on Monte Carlo to make use of deterministic and statistical knowledge of the uncertain and variable ocean environment to predict signal parameter pdf's. This talk

presents a statistical analysis of signals recorded during the SWellEX-96 measurement that took place in shallow water off the coast of California. The data consist of CW signals transmitted simultaneously from moving sources at two different depths and received at a horizontal array. The statistical analysis includes estimates of time-varying moments of signal parameters and coherence of signal amplitude, frequency, and phase. [Work supported by ONR Undersea Signal Processing Code 321US.]

4pSP7. Environmental and signal models for acoustic models for acoustic propagation through a variable ocean. Richard Lee Culver and H. John Camin (Appl. Res. Lab. and Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804, rlc5@psu.edu)

Acoustic signals propagating through the ocean are refracted, scattered, and attenuated by the ocean volume and boundaries. Although much about the ocean environment and its effect on acoustic propagation is understood, there is always variability associated with both environment parameters and the received signal, and uncertainty in our knowledge of both. Environment and signal models (e.g., sound speed, bottom loss, and received level) often include only mean values, but they can include the higher order moments that parametrize variability. Interest in relating environmental and signal models and variability is motivated by recent efforts to capture ocean environmental uncertainty that have enhanced understanding and models of environmental parameters, and by advances in ocean acoustic signal processing that indicate how to exploit knowledge of signal variability in order to improve sonar performance. These two developments can be connected if the relationship between environmental and signal parameters (or models) is understood. We show that when a signal prediction must reflect realistic, environment-driven variability over a particular space-time interval, the environmental models used to construct the signal model must contain the environmental variability present during this same interval. [Work supported by ONR Undersea Signal Processing Code 321US.]

4pSP8. Classification of active sonar detections with class specific method. Charles F. Gaumond and Ralph N. Baer (Acoust. Div. Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

The use of the Class Specific Method (CSM) [P. M. Baggenstoss, Trans. Sig. Proc. **47**, 3428–3432 (1999)] is demonstrated with numerically simulated data. The two models of a cylindrical shell and a fish school are shown to generate sufficient statistics with appropriate assumptions. The common reference signal for the CSM is Gaussian white noise. An interpretation of the resulting conditional probability density functions is given, which leads to a numerical regularization. The resulting implementation of the CSM is demonstrated using numerically modeled data of variable bandwidth with variable amounts of additive Gaussian white noise. This approach enables the evaluation of bandwidth requirements for sufficient classification performance. The method is shown to be generalizable to include effects of propagation. [Work supported by the Office of Naval Research.]

4pSP9. Statistical characterization of very low frequency communication channels at ocean basin-scales. John Spiesberger (Dept. Earth and Env. Sci., Univ. of Pennsylvania, Philadelphia, PA 19104-6316, johnsr@sas.upenn.edu) and Dale Green (Teledyne Benthos Inc., North Falmouth, MA 02556)

The bit rate for incoherent communication is estimated for low frequency sounds (75 Hz, 0.03 s resolution) transmitted 3115 km between a bottom-mounted source and a towed array. A model for short term fluctuations of sound agrees with measurements without tuning the model to the data. In particular, measured and modeled probability distributions for multipath coherence time and amplitude fluctuations are similar. Temporal evolution of the oceanographic component of the model is governed by the linear dispersion relation for internal gravity waves. The acoustic component of the model is based on the sound-speed insensitive parabolic approximation. This is the third attempt to use the model to predict short-term acoustic fluctuations at basin-scales. All three attempts yielded results similar to measurements. If further comparisons between the model and data are successful, the model could be used to design and implement low frequency communication systems. A simulation yields 4 bits per second for a Benthos modem using an incoherent communication scheme via frequency shift keying. [Work supported by the ONR.]

4pSP10. Some effects of spatial randomness along the length of the cochlear on its performance. Stephen Elliott, Emery Ku, and Ben Lineton (ISVR, Univ. of Southampton, Southampton, UK)

If the active cochlear is modelled as having a uniform distribution of parameters, the system is stable even for high gains in the cochlear amplifiers, resulting in very significant active enhancements of its response, of the order of 90 dB. In a real biological system, however, such uniform distributions will always be disrupted by random spatial variations, and a state space model of the cochlea has been developed that can be used to investigate the effects of such random variations. One effect is found to be that the cochlear amplifier gains have to be reduced to maintain stability, so that maximum active enhancements of the cochlear response is more typically 40 dB. The cochlear amplifier, however, is nonlinear as well as active, and unstable poles in the linear model generate instabilities that settle into limit cycle oscillations at discrete frequencies in the nonlinear model. These oscillations are thought to be the origin of spontaneous otoacoustic emissions, and have a significant effect on sound detection at low level. At higher sound pressure levels the nonlinearity in the cochlear amplifier suppresses such oscillations and the sound detection has a more uniform frequency response, although the amplification is reduced.

Session 5aMU

Musical Acoustics and Architectural Acoustics: Musical Requirements for Western and Non-Western Worship Spaces

Paul T. Calamia, Cochair
Rensselaer Polytech Inst., School of Architecture, 110 8th St., Troy, NY 12180

Ewart A. Wetherill, Cochair
28 Cove Rd., Alamedia, CA 94502-7416

Chair's Introduction—9:00

Invited Papers

9:05

5aMU1. Acoustical requirements for music in non-Western worship spaces. Paul Calamia and Jonas Braasch (Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

In this talk, an overview of different non-Western musical traditions for worship, and the particular architectural-acoustic requirements of these traditions, will be given. While the pipe organ, which plays a central role in the Christian and Jewish worship traditions, is rarely found in other religious services, the other two typical western worship instruments, choir and bells, serve in many other worship traditions as well. Brass instruments also are found in both Western and non-Western worship traditions, for example in the typical church brass ensembles with trumpets, trombones, horns, and tubas, and the long Tibetan temple trumpets. Percussion instruments, such as those in the Indonesian Gamelan Ensemble, are often an important element in non-Western rituals. However, aside from bells, they do not have a strong tradition in Western worship services, which is not surprising given the long reverberation times in most churches. The talk will focus on those worship traditions that are not addressed by other speakers in this session.

9:25

5aMU2. Music and acoustics for Catholic churches: history and modernity. Rendell R. Torres (Program in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, rendell.torres@gmail.com) and Kevin Galiè (Blessed John XXIII Natl. Seminary, 558 South Ave., Weston, MA 02493-2699)

The paradoxical challenge of providing crisp speech intelligibility and lush musical reverberance in Catholic liturgical spaces has long vexed acousticians, architects, and church building committees. Compounding the complexity is the variety of musical and liturgical styles, including ethereal Gregorian chant, post-Reformation strophic hymnody, and modern "praise music" (e.g., folk, gospel, and amplified pop music). In this paper we first survey official Catholic documents on music and sound in liturgy. Then we consider melodic structures and tempi of various kinds of liturgical music and relate them to acoustical decay parameters. Finally, we present a sample outline of music for a Catholic Mass and discuss attendant requirements from acoustical and musical perspectives.

9:45

5aMU3. Sound parameters in mosques. Wasim Asaad Orfali (Tech. Univ. Berlin, Inst. of Tech. Acoust., Einsteinufer 25, 10587 Berlin, Germany, orfaliwasim@hotmail.com)

Unlike auditoriums, there are no defined recommendations or rules for the acoustical parameters inside mosques. Most of the existing recommendations are developed for multipurpose halls, for opera or dramatic theaters, and for structures built for organ music. But regarding mosques the elaboration of general sound parameters requires a specific understanding of the acoustical and spiritual environment expected in such structures. Newly defined acoustical parameter values with regard to mosque volumes and types will be addressed here. New treatment rules for closed or courtyard structures are introduced. Especially so-called mosque volume dependent parameters are derived. They allow optimizing the secondary structure of a mosque by considering the primary structure determined by the architecture. Such parameters deal with optimal reverberation time or intelligibility considered for all praying modes at once. Target of the research is to get standard acoustic parameter values especially applicable for mosques.

10:05

5aMU4. Acoustics of music and voice in Jewish worship spaces. Mendel Kleiner (Chalmers Rm. Acoust. Group, Dept. of Appl. Acoust., Chalmers Univ. of Technology, 41296 Gothenburg, Sweden, mendel.kleiner@chalmers.se), David Klepper, and Yeshivat Beit Orot (Mt. of Olives, Jerusalem 97400, Israel)

Of the three monotheistic faiths, Judaism is the oldest, and one would expect synagogue architecture to predate mosque and church architecture. In fact, that is the case. The earlier synagogues actually were constructed while the Second Temple remained as the focal point for Jewish prayer, because local shrines and study halls developed during this period. Excavations indicate general simple rectangular halls of worship. Synagogue architecture developed in several styles and, of course, was influenced by both mosque and

5a FRI. AM

church architecture. The latter has been particularly true with non-Orthodox synagogues in Western countries, some of which have high-quality pipe organs and face the speech vs. music issue also faced by main-stream Protestant and Catholic churches. Meanwhile, Orthodox synagogues have become larger, and some Orthodox Rabbinic authorities have approved single-sided amplification, with current varied but without zero crossings. The most recent development in sound system technology, steerable pattern line sources, have presented the possibility of new and superior solutions to sound amplification in the worship spaces of all three faiths, supplementing the past solutions of central directional-horn-based systems, distributed systems of various types including pew-back, and delayed column loudspeakers, which may still have applications in some situations.

10:25

5aMU5. Acoustical design strategies for contemporary music in worship. Paul D. Henderson (Design 2020, 12082 Univ. City Blvd., Harrisburg, NC 28075)

Historically, acoustical design for traditional worship music has sought to naturally reinforce instruments and vocalists by carefully exploiting room reflections. However, the advent of "contemporary" worship and the widespread use of reinforced music have paralleled a clear change in the architecture, seating capacity, and expected acoustical signature of the worship sanctuary. In addition, many modern worship services employ a hybrid of musical styles within a single service, such as choral performances, orchestras, electronic music, and dramatic presentations. While some may consider a highly absorptive room to be an ideal listening environment, this approach neglects several key aspects of the contemporary worship experience, such as ensemble support for choirs as well as for the audience during congregational singing. Here, acoustical criteria are reviewed for the design of architectural elements within a contemporary worship space. Design strategies are presented for integrating effective acoustical elements within the framework of necessary technical systems. Case studies are presented of recent sanctuary designs of varying size to demonstrate the relationship between architectural techniques and the musical experience.

FRIDAY MORNING, 8 JUNE 2007 SALON B, 8:00 TO 10:05 A.M.

Session 5aPA

Physical Acoustics: Thermoacoustics

Richard Raspet, Chair
Univ. of Mississippi, National Center for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Chair's Introduction—8:00

Contributed Papers

8:05

5aPA1. Performance of a high frequency thermoacoustic cooler at high pressures. Husam El-Gendy and Orest G. Symko (Univ. of Utah, Dept. of Phys., 115 South 1400 East #201, Salt Lake City, UT 84112 orest@physics.utah.edu)

By raising the pressure of the working gas in a thermoacoustic cooler, the sound intensity level, at fixed electrical power input to the acoustic driver, is increased. In order to take advantage of this boost in acoustic intensity for cooling, significant adjustments need to be made inside the cooler. As the pressure is raised the thermal and viscous penetration depths are reduced; this requires an increase of stack filling factor accordingly. For the cooler presented here, operating at 4 kHz, the stack consists of a random arrangement of cotton wool, with fibers 10 microns in diameter. The stack is typically quite loose between the heat exchangers at each end. High levels of sound at the stack lead to volumetric forces on the fibers of the stack and viscous drag forces caused by acoustic streaming; both increase with the sound level. Moreover the heat capacity of the stack becomes insufficient to maintain the $T1=0$ condition at the interface with the working gas. Studies of such effects at high pressures will be presented.

8:20

5aPA2. Performance of annular high-frequency acoustic engines. Ivan A. Rodriguez and Orest G. Symko (Dept. of Phys., Univ. of Utah, 115 South 1400 East #201, Salt Lake City, UT 84112)

Thermoacoustic engines operating in the standing wave mode are limited in their efficiency by a high degree of irreversibility that is inherent in how they work. Better performance can be achieved by using traveling waves in the thermoacoustic device. This has led to the development of an annular high-frequency thermoacoustic prime mover consisting of a regenerator, which is a random stack in-between heat exchangers, inside an annular waveguide. Miniature devices were developed with operating frequencies in the range of 2–3 kHz. This corresponds to an average ring circumference of 11 cm for the 3-kHz device; the resonator bore is 6 mm. A similar device of 11 mm bore and length of 18 cm was also investigated; its resonant frequency is 2 kHz. Sound power was extracted by an impedance-matching side arm to the annular structure. The spectrum of the acoustic power did not have harmonics, indicative of all traveling wave components. The efficiency is high, being over 40% of Carnot. This type of device shows much promise as a high efficiency energy converter; it can be quite simply reduced in size for microcircuit applications.

8:35

5aPA3. High-resolution particle image velocimetry (PIV) measurements of the flow inside the stack-heat exchanger couple of a thermoacoustic refrigerator driven at high amplitude. Arganthael Berson and Philippe Blanc-Benon (Ctr. Acoustique–LMFA–Ecole Centrale de Lyon, 36 Av. Guy de Collongues, 69134 Ecully Cedex, France, philippe.blanc-benon@ec-lyon.fr)

The design of efficient heat-exchangers for thermoacoustic applications is one of the major challenges in the development of high-performance thermoacoustic coolers. With this prospect, an important step has been taken in this study with the characterization of the flow field around both stack and heat-exchangers, especially at high drive ratio when acoustics is nonlinear and edge effects become more important. High-resolution particle image velocimetry measurements are performed in a standing wave thermoacoustic refrigerator using a technique that has already been validated and successfully compared to numerical simulations Blanc-Benon *et al.*, C.R. Mécanique **331**, 17–24 (2003). High spatial resolution is achieved with an optical zoom mounted on a digital camera. Measurement areas down to 3×2.5 mm^2 are obtained. The resonator is driven at drive ratios up to 5%, with oscillating velocity amplitudes up to 15 m s^{-1} within the stack. As oscillating velocity amplitude increases, the vorticity layer developing behind each plate of a single stack becomes unstable and an asymmetric street of counter-rotating vortex pairs is observed. A Strouhal number based on the stack geometry is calculated. The consequences of vortex generation on the flow between a stack-heat exchanger couple are also investigated. Experimental data will be compared to numerical simulations from the literature.

8:50

5aPA4. Synchronization of small thermoacoustic oscillators. Brenna Gillman and Orest G. Symko (Dept. of Phys., Univ. of Utah, 115 South 1400 East #201, Salt Lake City, UT 84112)

Small high-frequency thermoacoustic oscillators can achieve high power densities in the conversion of heat to sound. In order to achieve high sound power output, it is advantageous to use arrays of such devices. Being a group of self-sustained oscillators, they switch on with a random phase at the threshold for oscillations. Moreover, small variations in the geometry of each device will affect the resonant frequency of each; the array will then not be in unison. This can be corrected by synchronizing them through global coupling; maximum power output will then be achieved. The strength of the coupling is assumed to be the ratio of mass of one thermoacoustic oscillator to the total mass of all devices and of the support to which they are mounted. By varying the strength of the coupling, the acoustic devices show synchronization, which for weak coupling is out-of-phase and at strong coupling is in-phase. Results are presented on the coupling of two prime movers operating near 2.5 kHz as well as for similar devices in an array consisting of five units. The synchronized behavior opens the field for large array systems and their applications.

9:05

5aPA5. Acoustic conversion of heat to sound at mid-audio frequencies. Bonnie McLaughlin and Orest Symko (Dept. of Phys., Univ. of Utah, 115 South 1400 East #201, Salt Lake City, UT 84112)

A thermoacoustic prime mover was developed for conversion of heat to sound that is then directly converted to electricity. The acoustic device consists of a 2.7-kHz quarter-wave resonator with a stack of random material between a hot heat exchanger and a cold heat exchanger. It is loaded by a cavity that couples the sound to a piezoelectric device for generation of electrical power. Optimization of this device for energy conversion was based on studies of heat injection, temperature difference threshold for onset of oscillation, heat flow in the device, quality factor Q of the resonator, response time to heat input, and sound power output. Parameters for optimization included different mesh sizes for the heat exchangers, given stack filling factors, and levels of positive feedback from the acoustic cavity. Response time to heat injection was lowered by coupling the heat source directly to the hot heat exchanger. Device efficiency was doubled

by reducing heat losses along the supporting structure of the stack. Temperature differences for oscillation were as low as 50 °C, and sound levels of 130 dB were achieved. Thus device performance was enhanced substantially by optimizing geometric factors.

9:20

5aPA6. Computational thermoacoustics using a lattice Boltzmann simulation. Carl Jensen (Dept. of Phys. and Astron. and the Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Computational fluid dynamics is being used to investigate the thermoacoustic properties of stack materials with irregular pore geometries. Available models for stack materials with irregular geometries (such as wools and foams) are based empiricutly on parameters of the material that must be acoustically measured. A direct simulation of the thermohydrodynamic acoustic flow using a thermal lattice Boltzmann method provides a model of the acoustics from the base parameters of the materials make-up and provides insight into the mechanisms affecting that behavior. A particular focus is being placed on fibrous stack materials because their simple geometry is less challenging to replicate on a computer. Work will be presented on modeling a simple straight pore for verification of the numerical procedure as well as progress towards fully modeling a fibrous material. [Work supported by U.S. Army Space & Missile Defense Command.]

9:35

5aPA7. Pressurization of high-frequency acoustic heat engines for power applications. Nick D. Webb and Orest G. Symko (Dept. of Phys., Univ. of Utah, 115 South 1400 East #201, Salt Lake City, UT 84112, orest@physics.utah.edu)

Acoustic heat engines show much promise for converting waste heat to electricity. Since most applications require high power levels, high frequency thermoacoustic engines can reach such performance by operating with pressurized working gas. Results on a 3-kHz prime mover, consisting of a quarter-wave resonator and a random stack between two heat exchangers, show that the sound intensity from such device is raised as the working gas is pressurized. At pressures up to approximately 10 atm, the increase in sound power is approximately linear to the increase in pressure, and thus is an effective way to increase the power output of thermoacoustic engines. Further increases in pressure lead to power output saturation because the filling factor needs to be adjusted for the change in thermal penetration depth of the gas. Pressurization also leads to lower Δ onset for oscillations, potentially opening up even more heat sources that can power a thermoacoustic engine. Pressurization promises to greatly increase the applications of acoustic engines to a variety of real world settings, providing a key source of renewable energy for the future.

9:50

5aPA8. Miniature acoustic prime mover operating at 10 kHz. Myra L. Flitcroft and Orest G. Symko (Dept. of Phys., Univ. of Utah, 115 South 1400 East #201, Salt Lake City, UT 84112)

A miniature thermoacoustic prime mover was developed for interfacing with electronics and microcircuits where it will be used for thermal management. It operates at approximately 10 kHz; heat is injected to the hot part of the device, the cold part is thermally anchored at room temperature. It consists of a quarter wave resonator of cylindrical geometry, 9 mm long. It has a stack of random porous material between two heat exchangers and it operates with air at one atmosphere. Preliminary results show that this oscillator produces a sound intensity at 110 dB, 1 m away from the open end of the resonator. This level can be raised substantially by careful alignment of the hot and cold parts of the resonator, and by reducing the heat losses along the stack and supporting structure between the hot and cold side of the stack. Based on the high operating frequency, a high power density is expected; such device can be used as the basic unit in an array of acoustic prime movers. This work provides a new functional miniature element that can lead the technology toward even smaller devices and well into the ultrasonic range.

Session 5aPP

Psychological and Physiological Acoustics: Detection, Masking, Temporal and Spectral Processing (Poster Session)

Bom Jun Kwon, Chair

Univ. of Utah, Communication Sciences and Disorders, 390 S. 1530 E., Suite 1201, Salt Lake City, UT 84112

Contributed Papers

All posters will be on display from 8:00 a.m. to 11:30 a.m. 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 9:45 a.m. and contributors of even-numbered papers will be at their posters from 9:45 a.m. to 11:30 a.m.

5aPP1. Analysis of observer-based measures of infant hearing. Raman Arora (Dept. of Elec. and Comput. Eng., Univ. of Wisconsin–Madison, Madison, WI 53705, ramanarora@wisc.edu) and Robert A. Lutfi (Univ. Wisconsin–Madison, Madison, WI 53705)

In the clinical evaluation of hearing, behavioral audiometry remains the gold standard. Yet behavioral measures are not easily obtained from those who stand to benefit most from early diagnosis of a hearing loss; infants and young children. In such cases behavioral assessment requires an observer-based method wherein a judge must decide whether the child has in some way responded to a sound. In the present paper observer-based measures are analyzed within the framework of a two-stage (infant-judge) detection model involving a minimum number of assumptions. Based on the results of this analysis, several recommendations are made for improving estimates of the hearing sensitivity of infants obtained from these measures. They include: (1) using multiple judges, (2) providing judges with prior knowledge of the stimulus on each trial, and (3) estimating the performance of judges at high signal levels. The last of these approaches, moreover, provides a means of empirically evaluating the former two and other procedural variations designed to improve estimates. [Work supported by NIDCD grant R01 DC01262-15.]

5aPP2. Comparison of behavioral and auditory brainstem measures of hearing loss in rats exposed to loud sound. Henry E. Heffner and Gimseong Koay (Dept. of Psych. MS 948, Univ. of Toledo, 2801 W. Bancroft St., Toledo, OH 43606, Henry.Heffner@utoledo.edu)

Behavioral thresholds for 400-ms pure-tone pulses (2/s, 10 ms rise-decay) were determined for monaural rats using the method of conditioned suppression. Auditory brainstem response (ABR) thresholds for 1-ms tone pulses (27.7/s, 0.5 ms rise-decay) were determined in the same animals for the same frequencies (2, 4, 8, 16, and 45 kHz). A hearing loss was then induced by exposing an isoflurane-anesthetized animal to a loud tone (110 or 120 dB) for 10 min. A behavioral pure-tone threshold was determined one hour after exposure followed immediately by an ABR threshold for the same frequency. The animals were retested on subsequent days until their thresholds stabilized. So far ($n = 10$), no simple relationship has been found between the ABR-estimated hearing loss and the actual hearing loss. Although the ABR tended to overestimate the hearing loss, the range of both over- and under-estimates was more than 25 dB. The difference between the two measures may be because of the different stimulus parameters used in the two procedures or because the ABR is a measure of neural synchrony as well as absolute sensitivity. Work is in progress to determine if a more reliable estimate of sensorineural hearing loss can be obtained from the ABR.

5aPP3. Absolute-detection results with a Bayesian adaptive tracking procedure using simulated and human subjects. Georg Stiefenhofer, William Woods, and Brent Edwards (Starkey Hearing Res. Ctr., 2150 Shattuck Ave. #408, Berkeley, CA 94704)

Adaptive tracking procedures are widely used in psychophysics. The common, fixed step-size, up-down procedures are known to be biased and yield convergence dependent upon the step-size used [e.g., B. W. Edwards and G. H. Wakefield, J. Acoust. Soc. Am. **83**, S17 (1988)], provide an estimate of only one parameter of the psychometric function, and in general make nonoptimal use of information (*a priori* and *a posteriori*) concerning a given subject's psychometric function. Recently introduced Bayesian adaptive procedures [e.g., L. L. Kontsevich and C. W. Tyler, Vision Res. **39**, 2729–2737 (1999)] make optimal use of all information in both the estimation of parameters of the psychometric function, as well as in stimulus placement via a minimum-entropy rule. No direct comparison of the up-down and Bayes procedures has been carried out, however. We present results of Monte Carlo simulations that demonstrate the Bayes procedure is as efficient, and often more efficient, than common up-down procedures, even when estimating more parameters than the up-down procedures. We will also present results from absolute-detection experiments with human subjects (with and without hearing loss) that demonstrate the Bayes procedures behavior outside the simulation environment. These experiments used *a priori* information derived from an audiogram database.

5aPP4. An examination of the relation between reaction time and loudness using perceptual weights. Lori J. Leibold (Dept. of Allied Health Sci., The Univ. of North Carolina at Chapel Hill, CB#7190, 301A S. Columbia St., Bondurant Hall, Chapel Hill, NC 27599, leibold@med.unc.edu) and Emily Buss (The Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599)

The relation between reaction time and loudness was examined by comparing the perceptual weights listeners assign to the individual components of a multi-tone complex during a reaction time task and a magnitude estimation task. Stimuli were 300-ms, five-tone complexes centered on 1000 Hz, with a logarithmic frequency spacing ratio of 1.586. Levels of the individual components of the complex were selected at random on each trial from a rectangular distribution with a range of 43–63 dB SPL. For the reaction time task, listeners initiated trials. The complex was presented following a random delay (500–4000 ms) to guard against anticipation effects. The subject was instructed to press a button on the response box as quickly as possible upon hearing a stimulus. All aspects of the magnitude estimation task were the same as described for the reaction time task, except that listeners gave absolute magnitude estimates for the loudness of the complex on each trial. Subjects performed 20 blocks of 50 trials each for the reaction time and the magnitude estimation tasks. Perceptual weights were computed by calculating the point-biserial correla-

tion between the level of each component and the subjects response. Results support the feasibility of obtaining reliable perceptual weights during both tasks.

5aPP5. Effects of masker phase on tone detection as a measure of cochlear compression in normal-hearing and hearing-impaired listeners. Van Summers, Matthew Makashay (Army Audiol. & Speech Ctr., Walter Reed Army Medical Ctr., Washington, DC 20307-5001), Marjorie Leek, and Michelle Molis (Portland VA Medical Ctr., Portland, OR 97207)

For listeners with normal hearing, Schroeder-phase harmonic complexes may vary in masking effectiveness by 20 dB or more depending on the temporal details of the stimuli. These masking differences are believed to be associated with curvature in cochlear phase response and compressive growth in the magnitude response for normal cochlear processing. Both curvature and compression may be greatly reduced in listeners with cochlear damage. In the current study, masker levels at masked threshold were determined for 1- and 4-kHz tones at a range of fixed signal levels using harmonic maskers varying in phase curvature. A Bekesy-tracking method, using a long-duration harmonic masker that slowly varied in temporal structure, was used to measure detection thresholds continuously over a range of curvatures. The masker producing the least effective masking and the overall change in masker effectiveness with phase structure were used to estimate phase curvature and degree of compression in cochlear processing for normal-hearing and hearing-impaired listeners. These estimates of cochlear compression will be related to auditory-filter bandwidths measures, also believed to be closely tied to peripheral compression. Further, thresholds from the fast Bekesy-tracking method will be compared with thresholds produced by more traditional, but time-consuming, forced-choice psychophysical methods.

5aPP6. The effect of the inter-target delay on informational and energetic masking of speech in younger and older adults. Liang Li, Ying Huang (Dept. of Psych., Peking Univ., Beijing, China), Qiang Huang, Xun Chen, Tianshu Qu, and Xihong Wu (Peking Univ., Beijing, China)

In a reverberant environment where several people talk simultaneously, perceptual grouping of the direct wave emanating from a talker with its reflections is essential for both distinguishing different talkers and identifying attended speech. Older-adult listeners often feel it difficult to understand attended speech in reverberant environments. In this study, the effect of changing the time interval between target speech and its simulated reflection on recognition of target speech was investigated in both younger and older adults with normal or near normal hearing. When the masker was competing speech, the intelligibility of target speech was markedly improved in younger participants as the inter-target delay was reduced from 64 to 0 ms, particularly when the signal-to-masker ratio was sufficiently low (-6 or -8 dB). In older participants, the release induced by decreasing the inter-target delay was marked but smaller than that in younger participants especially at long inter-target delays (16 or 32 ms). Moreover, reduced release of target speech occurred if speech masker was replaced by noise masker. The results suggest that there is an age-related decline of the ability to perceptually integrate target speech with its reflections. This ability is particularly important for releasing speech from informational masking in reverberant environments.

5aPP7. Reductions in gain from pre-masker stimulation at the signal frequency. Skyler Jennings and Elizabeth Strickland (Purdue Univ., West Lafayette, IN 47906)

Normal hearing individuals can detect a short duration signal at a lower signal-to-masker ratio when the signal is delayed from masker onset compared to near it. This is often referred to as the temporal effect. This effect is interesting because it may be related to changes in cochlear gain.

It has been proposed that gain is high at masker onset and decreases, in a frequency specific manner, through the course of the masker. If this hypothesis was true, it is expected that frequency selectivity also decreases through the course of the masker. Using a forward masking paradigm, the current study examines the hypothesis that frequency selectivity decreases with pre-masker stimulation at the signal frequency. Growth of masking (GOM) functions and psychophysical tuning curves (PTC) were measured for the signal using a short masker preceded by a long duration precursor set at the signal frequency or well below the signal frequency. Preliminary results reveal a relatively broader PTC and a steeper GOM function when the precursor was at the signal frequency and are consistent with the hypothesis that preceding stimulation at the signal frequency reduces gain. [Work supported by the Hazelton Fund.]

5aPP8. Spectral profile cues in comodulation masking release. Emily Buss (UNC Chapel Hill School of Medicine, 130 Mason Farm Rd., 1115 Bioinformatics Bldg., CB7070, Chapel HIll, NC 27510)

Previous work on spectral profile discrimination has shown that sensitivity is dependent on position of the signal in the standard complex, with a profile bowl shape demonstrated for components spanning 0.2–5 kHz [Green, Profile Analysis, 1988]. Dynamic spectral cues similar to those in profile analysis are present in comodulation masking release (CMR) stimuli, and the current study examined whether a similar bowl occurs for CMR. Maskers were 11 bands of noise, 15-Hz wide, logarithmically spaced 0.2–5 kHz, and maskers were either random or coherently modulated. The signal was either a pure tone at center frequency of one of these masker bands or the on-signal masker band itself. In the random conditions, thresholds were consistent across signal frequency for both signal types. In the coherent modulation conditions a bowl shape was obtained for the narrowband noise signal, but when the signal was a pure tone thresholds improved as a function of signal frequency and showed little evidence rising again at the highest signal frequencies tested. Data collected with the signal and on-signal masker presented contralateral to the flanking maskers indicate a reduced effect of signal type. Contribution of within- and across-channel cues will be discussed.

5aPP9. Effect of hearing loss on the ability to process synchronous and asynchronous cues for speech identification. Joseph W. Hall III, Emily Buss, and John H. Grose (Dept. Otolaryngol., UNC Chapel Hill, Chapel Hill, NC 27599-7070)

This study investigated the effect of sensori-neural hearing loss on the ability to identify speech in noise for vowel-consonant-vowel tokens that were unprocessed, modulated synchronously across frequency, or modulated asynchronously across frequency. The modulator was a 20-Hz raised square wave. One goal of the study was to determine whether hearing-impaired listeners have a deficit in the ability to integrate asynchronous spectral information in the perception of speech. Speech tokens were presented at a high, fixed sound level and the level of a speech-shaped noise was changed adaptively to estimate the masked speech identification threshold. The performance of the hearing-impaired listeners was generally worse than that of the normal-hearing listeners, but the impaired listeners showed particularly poor performance in the synchronous modulation condition. The results did not suggest that the hearing-impaired listeners have a general deficit in the ability to integrate asynchronous spectral information in the perception of speech. Results will be discussed in terms of common mechanisms that might account for poor speech identification performance by hearing-impaired listeners when either the masking noise or the speech is synchronously modulated.

5a FRI. AM

5aPP10. Gap detection in modulated noise: Across-frequency facilitation and interference. John H. Grose, Joseph W. Hall III, and Emily Buss (Dept. OHNS, Univ. of North Carolina, CB#7070, Chapel Hill, NC 27599-7070, jhg@med.unc.edu)

This study tested the hypothesis that a detection advantage for gaps in comodulated noise can be demonstrated in conditions of continuous noise and perceptually salient envelope fluctuations. Gap detection was measured using 25-Hz-wide bands of noise that were either Gaussian or were characterized as low-fluctuation noise or staccato (choppy) noise. The signal band was centered at 1000 Hz, and the flanking bands were centered at 444, 667, 1500, and 2250 Hz. Results from six observers confirmed that a gap detection advantage exists in continuous comodulated noise for Gaussian and staccato noise, but not for low-fluctuation noise. No gap detection advantage was found for a gated mode of presentation. Effects of temporal fringe and asynchrony suggested that the advantage was due to across-frequency processing and bore some resemblance to comodulation masking release (CMR). However, differences between CMR and the gap detection advantage were noted in terms of the effects of flanking band number and the absence of a detection advantage in gated conditions. This study demonstrates that, under some conditions, a detection advantage for gaps carried by a narrow band of noise can occur in the presence of comodulated flanking bands of noise. [Work supported by NIDCD R01-DC01507.]

5aPP11. The combined effects of reverberation and fluctuating noise on sentence intelligibility. Erwin L. J. George, Joost M. Festen, and Tammo Houtgast (Audiol. Dept., VU Univ. Medical Ctr., ENT, P.O. Box 7057, 1007 MB, Amsterdam, The Netherlands)

Listening conditions in everyday life typically include a combination of reverberation and nonstationary background noise. It is well known that sentence intelligibility is adversely affected by these factors. To assess their combined effects, a model is introduced that combines two models of speech perception, the Extended Speech Intelligibility Index (E-SII), and the Speech Transmission Index (STI). First, the effect of reverberation on nonstationary noise—reduction of modulations—is determined. The E-SII is then used to evaluate the effect of this modified nonstationary noise, while the STI is applied to quantify the effects of reverberation and noise on speech quality. To validate this model, speech reception thresholds (SRTs) were measured for ten normal-hearing listeners, under various combinations of nonstationary noise and artificially created reverberation. After taking the characteristics of the speech corpus into account, results show that the model accurately predicts SRTs in fluctuating noise and reverberation for normal-hearing listeners.

5aPP12. Rate versus time representation of high-frequency spectral notches in the peripheral auditory system: A computational model study. Enrique A. Lopez-Poveda, Ana Alves-Pinto, Almudena Eustaquio-Martin (Instituto de Neurociencias de Castilla y Len, Universidad de Salamanca, Avda. Alfonso X "El Sabio" s/n, 37007 Salamanca, Spain), and Alan R. Palmer (MRC Inst. of Hearing Res., Nottingham NG7 2RD, UK)

We investigated how high-frequency spectral notches are represented in the auditory nerve. Our approach consisted in modeling the paradoxical result of Alves-Pinto and Lopez-Poveda [J. Acoust. Soc. Am. **118**, 2548 (2005)] that discriminating between broadband noises with and without high-frequency spectral notches is most difficult at 70–80 dB SPL than at lower or higher intensities. We tested two possibilities: (a) that discrimination is based on the difference between the inner-hair-cell excitation patterns for the two stimuli, a representation related to the auditory nerve difference rate profile, and (b) that it is based on the difference inner-hair-cell modulation-rate pattern, a representation related to the difference in the temporal pattern of auditory nerve discharges. The simulations support the latter. This suggests that spectral features as high in frequency as 8 kHz may be encoded in the temporal pattern of auditory nerve responses despite the rapid decay of phase locking above 2 kHz. The simulations

also indicate that the improvement in spectral discrimination at high levels is associated to the saturation of the inner hair cell receptor potential.

5aPP13. The internal representation of high-frequency spectral notches. Ana Alves-Pinto and Enrique A. Lopez-Poveda (Instituto de Neurociencias de Castilla y Len, Universidad de Salamanca, Avda. Alfonso X El Sabio, 37007 Salamanca, Spain.)

To discriminate between broadband noises with and without a high-frequency spectral notch is more difficult at 70–80 dB SPL than at lower or higher levels [A. Alves-Pinto and E. Lopez-Poveda, J. Acoust. Soc. Am. **118**, 2548 (2005)]. One possible explanation is that the notch is less clearly represented internally at 70–80 dB SPL than at any other level. To test this hypothesis, forward-masking patterns were measured for flat-spectrum- and notch-noise maskers for masker levels of 50, 70, 80, and 90 dB SPL. To separate out the effects of peripheral compression and of the broadening of auditory filters with increasing level on the notch internal representation, the masking patterns were measured in two conditions: (1) varying the masker-probe intervals to achieve similar masked thresholds for different masker levels and (2) fixing the time interval at 2 ms. The depth of the spectral notch in the masking pattern decreased gradually with increasing level, mostly as a result of compression. This result is inconsistent with improvement in spectral discrimination above 80 dB SPL and hence undermines the common view that high-frequency spectral notches must be encoded in the rate-profile of auditory nerve fibers (see also the companion paper by Lopez-Poveda *et al.*).

5aPP14. Factors influencing the perception of non-familiar spectrally rotated speech syllables. Mahan Azadpour and Evan Balaban[a] (Cognit. Neurosci. Sector, SISSA (Intl. School for Adv. Studies), Via Beirut 2-4, 34014 Trieste, Italy, azadpour@sissa.it)

The effects of acoustic features and native phonology on the naïve perception of unfamiliar spectrally-rotated speech syllables were evaluated. Spectrally-rotated syllables were created by rotating the spectrum of naturally produced Italian CV syllables around 1500 Hz resulting in changes in the energy distribution of harmonics while preserving the temporal relationships of spectral trajectories. Italian-speaking participants with no previous spectrally-rotated speech experience were asked to describe spectrally-rotated syllables phonologically by typing onomatopoetic versions of their perceptions on a computer keyboard. The results, also used to estimate the phonological similarity of the rotated sounds, showed that the identification of rotated vowels was related to their location on the F1-F2-F3 map of native (unrotated) vowels. The discriminability of rotated sounds was examined using a separate, naïve subject group using an ABX procedure. Discrimination errors were positively correlated with the estimated phonological similarity between rotated sounds. For rotated vowels, discrimination errors were also related to their Euclidean distance in F1-F2-F3 space as revealed by regression analysis. We conclude that naïve perception of unfamiliar spectrally-rotated syllables is under the influence of native phonology, and that spectrally-rotated vowels are represented by their location in the formant space. [a] Also at Behavioral Neurosciences Program, McGill University, Montreal, Canada.

5aPP15. Numerical and experimental study of bone conducted sound and the occlusion effect. Lynn M. Brault, Sara Melamed, Charissa R. Lansing, Ron D. Chambers (Dept. of Speech and Hearing Sci., UIUC, 901 S. 6th St., Champaign, IL 61820), and William D. O'Brien (Bioacoustics Res. Lab., Urbana, IL 61801)

Numerical techniques are used to study bone conduction hearing mechanisms. The military has an interest in bone conducted sound as personnel continue to experience hearing loss despite availability of full hearing protection. Particularly they need to identify better hearing protection over the entire audio range. A joint numerical/experimental program has been undertaken to quantify bone-conducted sound, identify paths

along which sound propagates, and determine frequency-dependent characteristics of this conducted pressure. Specific cases in this work include coupling sound to the mastoid and measuring resultant pressures in the ear canal with and without earplugs. Canal pressure can be higher with ears blocked as plugs will trap conducted sound in a resonance mode, a phenomenon known as occlusion. In order to simulate this effect, digital images of a skull CT scan were used as the finite difference grid. Experimental results, based on behavioral and physiologic responses, were obtained from participants at UIUC to provide validation of the simulation. In both cases the occlusion effect is shown to be frequency dependent. The simulation software offers the possibility of evaluating the effect over a wide range of parameters such as earplug type, skull shape, and location of inputs. [Work supported by AFOSR FA9550-06-1-0128.]

5aPP16. A model with compression for the estimation of speech intelligibility in quiet. Johannes Lyzenga (Dept. of ENT/Audiol., VU Univ. Medical Ctr., De Boelelaan 1117, 1185AT, Amsterdam, The Netherlands) and Koenraad S. Rhebergen (Academic Medical Ctr., Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands)

The Speech Intelligibility Index (SII) is an often used calculation method for estimating the proportion of speech information available in noise. For speech reception thresholds (SRTs), measured in normal-hearing listeners using various types of stationary noise, this model predicts a fairly constant speech proportion (about 0.30) necessary for sentence intelligibility. However, when the SII model is applied for SRTs in quiet, the estimated speech proportions are often lower, and show a larger inter-subject variability, than found for speech in noise near the normal speech level (65 dB SPL). The present model attempts to alleviate this problem by including cochlear compression. It is based on a loudness model for normal-hearing and hearing-impaired listeners of Moore and

Glasberg [Hear. Res. **188**, 70–88 (2004)]. It estimates internal excitation levels of the speech and then calculates the proportion of speech above threshold using similar spectral weighting as used in the standard SII. The present model and the standard SII were used to predict SRTs in quiet for both normal-hearing and hearing-impaired listeners. The present model predicted a set of data for three listener types (normal hearing, noise-induced hearing loss, and age-induced hearing loss) with less variability than the standard SII.

5aPP17. The significance of offset transients in listener identification of pipe organ stops. Alastair C. Disley and David M. Howard (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK)

Onset transients have been demonstrated to have a significant effect on listener identification of pipe organ stops and have been widely studied (Disley, A. C., and Howard, D. M., (2006). Onset transient significance in listener identification of pipe organ stops, J. Acoust. Soc. Am. **119**, 3333, abstract only). Offset transients in contrast are frequently ignored, but the authors' previous research suggests that for certain stops, in particular reed stops, the offset transient is a significant cue in correct listener identification. Eight samples of individual pipes, covering most common stop types, were presented to trained organists in a listening test. Each sample was presented with and without its offset transient, the transient being replaced by an amplitude ramp in the latter case. To avoid an obvious curtailment of a reverberant tail, all samples were recorded within a closed swell box, and then had reverberation added using impulse responses of real churches to preserve the impression of a real and typical listening environment. The results demonstrate that offset transients are important in listener identification of some stops, but the significance varies between different classes of stop.

FRIDAY MORNING, 8 JUNE 2007

ALPINE WEST, 8:30 TO 11:45 A.M.

Session 5aSC

Speech Communication: Voice and Prosody (Lecture/Poster Session)

Yi Xu, Chair

University College London, Phonetcs and Linguistics, Wolfson House, 4 Stephenson Way, London, NW1 2HE, UK

Chair's Introduction—8:30

Contributed Papers

8:35

5aSC1. Interaction of word stress, focus, and sentence type in English. Fang Liu (Dept. of Linguist., Univ. of Chicago, 1010 E. 59th St., Chicago, IL 60637, liufang@uchicago.edu) and Yi Xu (Univ. College London, London, NW1 2HE, UK)

This study investigates prosodic differences between statements and declarative questions in American English by examining their interaction with focus, as manifested in the local and global pitch range variations, and in the underlying pitch targets of the stressed syllables in the pre-, post-, and on-focus content words. Five native speakers read 24 sentences eight times. Results of F0 analyses indicate that focus has no effect on the prefocus region for either statements or questions. In the on-focus region, the pitch range of the stressed syllable is raised in statements, but lowered in questions. The postfocus pitch range is compressed and lowered in statements, but compressed and raised in questions. Furthermore, the pitch target of the stressed syllable in a content word is high or falling in statements and low or rising in questions, depending on the focus condition and on the word's stress and syllabic pattern. Comparisons of duration and

spectral balance in each statement-question pair yielded no significant difference. These results suggest that a particular combination of word stress, focus, and sentence type in an English utterance largely determines its local and global pitch contours. Only after these communicative functions are identified can the associated forms be explained.

8:50

5aSC2. How often is maximum speed of articulation approached in speech? Yi Xu (Univ. College London, Wolfson House, 4 Stephenson Way, London NW1 2HE)

According to the principle of economy of effort, dynamic extremes such as maximum speed of articulatory movement would be avoided as much as possible. Recent findings about F0 dynamics suggest, however, that maximum speed of pitch change is actually often approached in the production of tone and intonation. Here we investigate the possibility that segmental production also often involves motor movements that approach the maximum speed of articulation. Native speakers of American English

3199 J. Acoust. Soc. Am., Vol. 121, No. 5, Pt. 2, May 2007

153rd Meeting: Acoustical Society of America **3199**

5a FRI. AM

were asked to produce repetitive syllable sequences such as yayayayaya and wiwiwiwiwi as fast as possible by imitating the same sequences that had been artificially accelerated through resynthesis. They were also asked to produce meaningful sentences containing the same syllables at normal and fast speaking rates. The speed of articulation in the syllable sequences and in the sentences are compared in terms of peak velocity of formant movement as a function of movement amplitude. If peak velocity is comparable across the conditions, it is an indication that speech production is operated near the performance ceiling, which would imply that much of the contextual variability in speech, especially that related to undershoot, has more to do with dynamic constraints of articulation than with economy of effort.

9:05

5aSC3. The effect of voicing, pitch, and vocal tract length on the recognition of concurrent speech. Martin D. Vestergaard, Nicholas R. C. Fyson, and Roy D. Patterson (CNBH, Univ. of Cambridge, CB2 3EG, UK, mdv23@cam.ac.uk)

In natural speech, the pitch of the voice can be varied by the speaker, but they cannot vary their vocal tract length (VTL), so VTL would seem to be a logical choice of tracking variable for listening in multispeaker environments. Experiments with concurrent speech have shown that differences in pitch and VTL make it possible for listeners to attend to a particular speaker in a multi-speaker environment, but the relative contribution of pitch and VTL is not clear. A series of experiments was conducted to investigate the extent to which differences in vocal characteristics contribute to the recognition of concurrent speech, using syllables with matched temporal envelopes to preclude glimpsing. Syllable pairs were synthesized with a vocoder to simulate different speakers. We measured the recognition of whispered and voiced syllables masked by syllables that varied in voicing, pitch, and/or VTL. The results show that differences in any of the vocal characteristics improve recognition performance. Voiced speech was more recognizable than whispered speech at the same SNR when it was the target speaker. When the distracter was whispered speech, it produced lower recognition scores than voiced speech. [Research supported by the UK-MRC (G0500221, G9900369).]

9:20

5aSC4. Parsing of ambiguous phrases: Single versus multiple prosodic cues. Erin M. Good (Univ. of Arizona, P.O. Box 210028, Tucson, AZ 85721-0028, emgood@email.arizona.edu)

In English, two prosodic contours that differ acoustically in their pitch and the location of the pauses involved can be used to disambiguate two-word phrases. The list contour, which has a fall-rise in pitch over the first word and a pause (approximately 22 ms) between the words is distinguished from the compound contour, which has a rise-fall in the pitch over both words and no pause. In this experiment target words recorded with relatively flat pitch were spliced together to make phrases with either 10 or 40 ms between each word. Participants treated items containing the 10-ms pauses as if they were compounds and items with 40-ms pauses as lists even without a pitch cue. This was true for lexically ambiguous phrases

where either prosodic contour was appropriate and for nonsense phrases where the prosodic contour alone made it possible to select the correct answer. There was no statistical difference between the responses to items with just the pause cue and to items with the full prosodic contour. While many prosodic phenomena have multiple cues, the data suggest that in some cases a single cue can be used in place of multiple cues to the same effect.

9:35

5aSC5. A prototypical pitch pattern of unaccented Japanese words. Irina Shport (Linguist. Dept., 1290 Univ. of Oregon, Eugene, OR)

Unaccented words are described as having a rising pitch pattern in Tokyo Japanese. The production and perception of $F0$ patterns in three-mora unaccented words that varied in segmental type of the word-medial mora were investigated. In production ($N = 864$), $F0$ rose across words having a short vowel in the initial syllable (V-type), and $F0$ was flat across words having a long vowel or nasal coda in the initial syllable (V:/VN-type), $p < .01$. In perception ($N = 384$), 70% of words with the V-type first syllable were correctly discriminated from accented words, whereas 56% of words with the V:/VN-type first syllable were correctly discriminated. Pairwise comparisons between word types revealed that the accentual pattern of CVC:V words was discriminated significantly better (by 24%) than that of CVNCV words, $p < .01$. These results indicated that unaccented words with rising, as opposed to flat $F0$, were more accurately discriminated from accented words. This might be due to a non-prototypical nature of flat $F0$ patterns in three-mora unaccented words. In other words, the mental representation of unaccented words may be characterized by a rising pitch pattern.

9:50

5aSC6. Spectral emphasis as a stress cue in the Latinamerican Spanish. Alejandro C. Renato (Inst. Genius of Technol.. Av. Aca 875 Bloco E Distrito Industrial Manaos AM Brasil)

Several studies have reported that spectral emphasis is a reliable correlate of stress, e.g., [Sluijter and van Heuven, J. Acoust. Soc. Am. **100**(4), 2471–2485] that is to say, harmonics situated above 0.5 kHz present more prominence in a stressed vowel. The spectral balance or spectral emphasis was then examined as a cue related to loudness or intensity. The present work analyzes the relationship between the described phenomenon and the vowel pitch, in order to relate the contribution of higher harmonics with the probable perceived pitch and evaluate their actual weight. The virtual pitch theory [Terhardt, Stoll and Seewann, J. Acoust. Soc. Am. **71**(3), 679–688] provides a way to understand this relationship. To achieve this purpose a speech production study was done with native latino-american spanish speakers. The corpus consisted of minimal stress pairs of bi and trisyllabic lexical items with and without stress in a carrier sentence. Analysis was developed in two steps. In the first step were obtained the values in frequency and sound pressure level in decibels for a frame of each harmonics through a tracker algorithm. In a second step, the virtual and spectral pitch was extracted from the first step data, in accordance with pitch virtual theory. Results show an increase of weight of spectral pitch in frequences superior to 0.5 kHz.

10:05–10:15   Break

All posters will be on display and all authors will be at their posters from 10:15 a.m. to 11:45 a.m.

5aSC7. Effects of vocal pitch level and perturbation magnitude on the pitch-shift reflex. Hanjun Liu and Charles R. Larson (Dept. of Commun. Sci. and Disord., Northwestern Univ., Evanston, IL 60208)

The purpose of the present study was to investigate the responsiveness of the pitch-shift reflex to small magnitude stimuli, vocal level, and pitch-shift direction. English speakers received pitch-shifted voice feedback (10, 20, 30, 40, and 50 cents, 200 ms duration) during vowel phonation at a high pitch and a low pitch level. Mean response magnitude increased as a function of stimulus magnitude, but when expressed as a percent of stimulus magnitude, declined from 100% with 10 to 37% with 50 cent stimuli. Response magnitudes were larger and latencies were shorter with a high

pitch level (15.8 cents; 130 ms) than a low pitch level (12.8 cents; 152 ms). Similar results were found with downward stimuli compared to upward stimuli (15.7 cents versus 13.1 cents; 126 ms versus 155 ms). Data from the present study demonstrate that vocal response magnitudes are equal to small perturbation magnitudes, and they are larger and faster with a high-pitched voice and downward directed pitch stimuli. These results suggest that the audio-vocal system is optimally suited for compensating for small pitch rather than larger perturbations. Data also suggest the sensitivity of the audio-vocal system to voice perturbation may vary with pitch level.

5aSC8. Voice F0 responses elicited by perturbations in pitch of auditory feedback in individuals that stutter and controls.

Jay J. Bauer, Carol Hubbard Seery, Rachelle LaBonte, and Lauren Ruhnke (Dept. of Commun. Sci. & Disord., Univ. of Wisconsin—Milwaukee, P.O. Box 413, Milwaukee, WI 53201)

This study aims to increase our understanding of the underlying mechanisms of vocal control in frequency-altered conditions of auditory feedback for people who do and who do not stutter by assessing the vocal responses to perturbations in pitch of auditory feedback. Past research has shown that the speech of people who stutter improves during frequency-altered feedback, but the mechanisms responsible for this fluency remain unclear. Typically, brief modulations in pitch of voice auditory feedback lead to short-latency corrective voice F0 responses during sustained phonations and inflected speech in nonstuttering individuals. However, data are lacking regarding audio-vocal control mechanisms in individuals that stutter. Brief upward $(+)$ and downward $(-)$ perturbations in pitch (50 or 600 cents in magnitude) lasting 200 ms in duration were introduced intermittently into vocalizing subjects auditory feedback. To date, $N=3$ moderate to severe developmental stutterers and $N=3$ age and gender matched control nonstutterers have been tested using the pitch-shift paradigm. Preliminary analyses indicate that compensatory voice F0 responses during sustained vowels are less prevalent, more variable, and slower in latency in individuals that stutter compared to age and gender match controls. However, subject recruitment and analyses related to inflected speech productions are still ongoing.

5aSC9. Information conveyed by voices.

Bruce R. Gerratt and Jody Kreiman (Div. Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794, bgerratt@ucla.edu)

Two approaches for dealing with interrater variability in voice quality judgments have recently been proposed. The first [Shrivastav et al., J. Speech Lang. Hear. Res. **48**, 323–335 (2005)] advocates reducing rating errors after data collection by averaging many ratings of each voice by each rater and then standardizing the resulting means. The second [Gerratt and Kreiman, J. Acoust. Soc. Am. **110**, 2560–2566 (2001)] controls rating error at the source by applying a method-of-adjustment task, which controls variability due to attention, variable internal standards, etc. Using this second method, we accounted for 87% of the variance in listeners perception of noise level for 40 voices. This low level of residual variance (errors) raises the following question: Does averaging ratings reduce the information within listeners responses along with measurement errors? To address this question, 20 listeners provided 10 breathiness ratings for 40 voices. The likelihood of exact listener agreement was calculated and compared to data from the method-of-adjustment task. We also compared amounts of variance accounted for by differences between voices in quality (a metric of information captured). We hypothesized that the averaging and standardizing approach would produce both lower interrater agreement and lower amounts of variance accounted for than the method of adjustment technique. [Work supported by the NIH/NIDCD.]

5aSC10. The voice use profile; illustrating actual voice use from long term monitoring from the National Center for Voice and Speech voice dosimeter.

Eric Hunter, Ingo Titze (Natl. Ctr. for Voice and Speech, The Denver Ctr. for the Performing Arts, Denver, CO), and Ingo Titze (Univ. of Iowa, Iowa City, IA)

The Voice Range Profile (VRP) is a map of the dynamic and fundamental frequency ranges of a voice and is often used by both the voice clinicians and coaches to show improvements in therapy and/or training. However, although the VRP may accurately capture the extreme range of the voice, it does not represent actual patient/student voice use which is where the therapy and training hope to improve. Using the National Center for Voice and Speech voice dosimeter, the actual voice use of 60 teachers was captured over two weeks per subject (average of 160 hours of recording per subject). The result is an entirely new method of examining voice; the Voice Use Profile (VUP). The VUP enhances the VRPs 2-D representation of vocal frequency and intensity by adding a third dimension, that of normalized incidences of occurrences. Using the VUP, intensity-frequency area where a voice user spend most of his/her time can be quantified and tracked, allowing for metrics for patient improvements. The presentation of this abstract will illustrate the VUP of several subjects showing how the addition of this third-dimension gives a detailed look of the topography of the subjects voice use through computer animation.

5aSC11. Voicing and silence periods in daily and weekly vocalizations of teachers.

Ingo Titze (Dept. of Speech Pathol. and Audiol., Univ. of Iowa, Iowa City, IA 52242), Eric Hunter (The Denver Ctr. for the Performing Arts, Denver, CO), and Jan Vec (Groningen Voice Res., Univ. of Groningen Medical Ctr., Groningen, The Netherlands)

The National Center for Voice and Speech data bank on voice dosimetry was used to study the distributions of continuous voicing periods and silence periods in 31 teachers over the duration of two weeks. Recordings were made during all awake hours of the day. Voicing periods were grouped into half-decades, ranging from 0.0316 0.10 s for the shortest periods of phonation to 31.6 100 s for the longest periods of phonation. Silence periods were grouped into similar half-decades, but ranged up to periods of several hours. On average, the teachers had 1,800 occurrences of voicing (onset followed by offset) per hour at work and 1,200 occurrences per hour while not at work. Voicing occurred 23% of the total time at work, dropping to 13% during off-work hours and 12% on weekends. The greatest accumulation of voicing occurred in the 0.316 1.0 s voicing periods, whereas the greatest accumulation of silence occurred in the 3 10 s silence periods. The study begins to lay the groundwork for understanding vocal fatigue in terms of repetitive motion and collision of tissue, as well as recovery from such mechanical stress.

5aSC12. Computational study of glottal flow aerodynamics.

L. Joel Peltier and Michael Krane (ARL, Penn State Univ., P.O. Box 30, State College, PA 16804)

We present model simulations for flow through an idealized, time-varying glottis as a building block step to a fully coupled fluid-structure interaction model. The CFD code used is AcuSolve, a finite-element flow solver, second-order accurate in space and time, from AcuSim Corp. of Mountain View, CA. Relative mesh motion in the glottis is achieved using a sliding mesh technique. The code supports highly skewed cells, allowing glottal closure. The geometric configuration and flow conditions of the computed flow match the rigid vocal fold model experiments ($4500 < Re < 11,000$, $0.01 < f^* < 0.06$) reported by Krane et al. [J. Acoust. Soc. Am. (2007)]. Details of the pressure and velocity fields and how they evolve with opening and closing of the glottis are presented, extending the information contained in the experimental database.

5aSC13. Simulations of vocal fold movement and aerodynamics using the Uintah Computational Framework.

Comer Duncan (Dept. of Phys. and Astron., Bowling Green State Univ., Bowling Green, OH 43403), Ronald Scherer (Green State Univ., Bowling Green, OH 43403), James Guilkey, and Todd Harman (Univ. of Utah, Salt Lake City, UT 84112)

This study applies a tightly coupled fluid-structure interaction algorithm to the modeling of phonation. The Uintah Computational Framework models vocal fold material using the Material Point Method (MPM), which permits arbitrarily large material displacements and multiple materials characterizing the vocal fold properties. The air is modeled using a compressible Navier-Stokes solver, the Implicit Continuous-fluid Eulerian (ICE) method developed at Los Alamos National Laboratory by B. A. Kashiwa. The MPM and ICE methods are coupled together to generate a unique simulation tool. Preliminary simulations are shown of 2-D model vocal folds interacting with prescribed transglottal pressures between 100 Pa and 800 Pa illustrating the intrinsic coupling between the vocal folds

and the air. Results are presented showing how the new simulation scheme characterizes the materials and how the aerodynamics that results displays the essential characteristics of glottal flow. The next steps of incorporating three-dimensionality and acoustics will be discussed. The present simulations set the stage for a realistic computational glottis and for eventual modeling of the effects of vocal fold pathologies on the acoustical output. [Work supported at Bowling Green State University in part by the National Institutes of Health and at the University of Utah by the Department of Energy.]

5aSC14. Prosodic features of Mandarin utterance final particles. Bin Li (Dept. of Chinese, Translation & Linguist., City Univ. of Hong Kong, 83 Tat Chee Ave., Kowloon Tong, Hong Kong, China)

Utterance final particles (UFPs) in Mandarin are generally regarded as discourse markers. The focus of this paper is two most frequently used UFPs: NE and A, which share overlapping discourse functions and may appear in identical utterances. From traditional discourse perspectives, NE has been described as functioning in a variety of ways, such as emphasizing, reminding, and retorting. Functions of A may include explaining, reminding, and affirming. These characteristics, more idiomatic than grammatical, pose difficulty in language processing and second language acquisition. This study proposes an acoustic approach to the study of these UFPs, and hypothesizes a trade-off relationship between the prosody of an utterance and a UFP. Our first aim is to investigate the prosodic features of utterances with and without particles, testing the hypothesis on a trade-off relationship between a UPF and its context. Our second aim is to compare across particles NE and A in search of categorizing acoustic correlates, as the two are among the most confused particles by Mandarin learners.

FRIDAY MORNING, 8 JUNE 2007

CANYON ROOM C, 8:00 TO 11:00 A.M.

Session 5aUW

Underwater Acoustics: Detection and Array Processing

Ying-Tsong Lin, Cochair

Woods Hole Oceanographic Inst., Applied Ocean Physics and Engineering, 210 Bigelow Bldg., MS 11, Woods Hole, MA 02543

Purnima Ratilal, Cochair

Northeastern Univ., Electrical and Computer Engineering, 409 Dana Research Center, Boston, MA 02115

Chair's Introduction—8:00

Contributed Papers

8:05

5aUW1. Numerical prediction of coherent integration time at 75 Hz, 0.03 s temporal resolution at 3250 km. John Spiesberger (Dept. Earth and Env. Sci., Univ. of Pennsylvania, Philadelphia, PA 19104-6316, johnsr@sas.upenn.edu)

Coherence time of sound is modeled for a 3250 km section in the Pacific at 75 Hz, 0.03 s resolution. The model is based on the temporal evolution of a standard spectrum of internal gravity waves and an accurate approximation for the acoustic wave equation. Modeled coherence time extends up to about 40 min. The probability is 0.6 that the signal-to-noise ratio first decreases at times of 13 min or less. Afterward, the s/n usually increases, reaching a peak near 40 min. The 40 min time is longer than experimentally reported coherence times of 12.7 min [J. Acoust. Soc. Am., **105**, 3185–3201 (1999)]. The experimental value is similar to the modeled time of 13 min or less at which the probability is 0.6 for finding the s/n ratio to decrease for the first time. The experimental value of 12.7 min was found using a single Doppler correction speed yielding maximum output of the impulse response. Analysis of accelerations of the source and receiver indicates a variable Doppler correction scheme should be used to estimate coherence time. The experimental procedure for assuming a constant Doppler speed correction appears to be invalid. [Work supported by the ONR.]

8:20

5aUW2. Transverse coherence in a winter ocean. Peter C. Mignerey (Acoust. Div. 7120, Naval Res. Lab., Washington, DC 20375-5350)

In December 2003 an ocean acoustics experiment was conducted in 75 m of water on the New Jersey Shelf in a nearly isothermal ocean. Sources broadcast 300 and 500 Hz narrowband signals at 15 dB SNR along a 20 km range to a bottomed horizontal line array oriented transverse to the acoustic propagation path. Estimates of the mutual coherence function were obtained by ensemble averaging 137 748 samples obtained over a period of 13 days, and compared to a canonical exponential form. It was found that the data were not properly described by an exponential. Rather, the data comprise two regimes: one with an exponential decay rate between 0.8–1.2 and coherence length of 12 wavelengths, the other with an exponential decay rate of 0.3 and coherence length of 27–166 wavelengths. The changeover between the two regimes occurs at hydrophone separations of 4–9 wavelengths. A scintillation index near one indicates that the acoustic fluctuations are dominated by phase variability. It is concluded that these data are not well-described by a simple exponential form and that work is necessary to develop an acoustic fluctuation theory appropriate for this environment. [Work supported by the Office of Naval Research.]

5aUW3. Holographic and conventional bistatic synthetic aperture sonar imaging of cylinders: Comparison and ray interpretation. Kyungmin Baik, Christopher Dudley, and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

When synthetic aperture sonar is used to image elastic targets, subtle features can be present in the images associated with the dynamical response of the target being viewed. In an effort to improve the understanding of such responses as well as to explore alternative image processing methods, a laboratory-based system was developed in which targets were illuminated by a transient acoustic source and bistatic responses were recorded by scanning a hydrophone along a horizontal line. Images were constructed using a relatively conventional bistatic SAS algorithm and were compared with images based on supersonic holography. The holographic method is a simplification of one previously used to view the time evolution of a target's response [B. T. Hefner and P. L. Marston, "Acoustics Research Letters Online" **2**, 55–60 (2001)]. In the holographic method, the time evolution of the cross-range image was used to construct a two-dimensional image with an effective range coordinate. Various features for cylindrical targets were interpreted using quantitative ray theory. This includes contributions from guided surface elastic waves as well as transmitted-wave features. [Work supported by the Office of Naval Research.]

5aUW4. Impact of a shelfbreak front on array processing in shallow water. Ying-Tsong Lin, Arthur E. Newhall, and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

In a coastal ocean shelfbreak front region, underwater sound can horizontally refract in the offshore direction due to the sloping bottom and reflect/refract back shoreward due to the ocean front. To study this interesting phenomenon, an analytic solution for an ideal wedge with a frontal interface was introduced in our previous paper [Lin *et al.*, J. Acoust. Soc. Am. **120**, 3182 (2006)]. This analytic solution suggests that the sound will also tend to propagate along the front, and that so-called "whispering gallery" modes can be observed. In this paper, the consequence of this whispering gallery effect on array processing is discussed. Specifically, the influence of an ideal ocean front on the array gains is studied. Since the sound reflecting from the ocean front forms highly correlated beams, the array gain of a horizontal hydrophone array will be increased. In addition, the application of modal beamforming techniques to resolve the whispering gallery modes propagating along the front is demonstrated. Our research goal for this study is to pursue practical systems to locate an inshore, long-distance sound source and also to monitor the changes of ocean fronts.

5aUW5. The effect of non-Gaussian noise on detection of sinusoids in the ocean. Colin W. Jemmott, R. Lee Culver, Jeffrey A. Ballard, and Leon H. Sibul (The Penn State Univ., P.O. Box 30, State College, PA 16804)

Passive sonar detection algorithms are often derived under the assumption that the signals of interest are embedded in additive white Gaussian noise (AWGN). However, measurements of low-frequency ambient noise in the ocean are often poorly modeled by this distribution. Detectors derived under the AWGN assumption will suffer lower performance in the presence of non-Gaussian noise. In this paper, ambient noise is modeled using the generalized Gaussian distribution, a three-parameter distribution that can be adjusted so as to match many sources of low-frequency ambient ocean noise. The signal model is sinusoidal with known frequency and amplitude but random phase. Unfortunately, the resulting likelihood ratio is not mathematically tractable for this signal and noise model, so several approximations to the optimal detector are investigated. Using Monte Carlo simulation, the performance of the detectors is compared for a variety of signal and noise parameters.

5aUW6. Coherent-broadband, white noise constraint, matched-field processing. Claire Debever and William A. Kuperman (Marine Physical Lab, Scripps Inst. of Oceanogr., 8820 Shellback Way, NTV Bldg., Rm. 446, La Jolla, CA 92093, cdebever@ucsd.edu)

Coherent-broadband matched-field processing takes advantage of the temporal complexity of the signal and, therefore, offers an increased gain over narrowband processing. However, the problem of mismatch is aggravated in adaptive versions such as the MVDR. One constraint is not enough to achieve robustness in such high-dimensional spaces. Some coherent algorithms also require a large number of snapshots, leading to ill-conditioned cross-spectral density matrices in the context of the required MVDR matrix inversions. The white noise constraint algorithm developed by Cox *et al.* [IEEE Trans. on Acoustics, Speech, and Sig. Proc. **35**, 1365–1376 (1987)], and typically used for narrowband processing, is applied to some previously derived broadband processors to enhance their robustness. The broadband white noise constraint theory is presented and validated through simulation and experimental data. [Work supported by ONR.]

5aUW7. Multi-channel detection of weak narrow-band signals in the presence of Gaussian and impulsive noise. Berke M. Gur and Christopher Niezrecki (Dept. of Mech. Eng., Univ. of Massachusetts-Lowell, One University Ave., Lowell, MA 01854, berke_gur@student.uml.edu)

Signal detection in the presence of noise is a commonly encountered problem in underwater acoustic signal processing and arises in many applications involving active and passive sonar. The detection problem becomes especially challenging if the signal to be detected is weak compared to background noise. In this work, the problem of detecting weak narrow-band bio-acoustic signals in noisy underwater acoustic environments is addressed. In contrast to most previous work, multi-channel setups are also considered and improvements in detector performances using multi-channel setups are investigated. Both Gaussian and non-Gaussian impulsive noise models are studied. Several parametric and nonparametric detectors are developed and implemented. The detection performance and range of these detectors in detecting manatee vocalizations buried in actual underwater acoustic noise recordings dominated by boat noise and snapping shrimp crackles are evaluated. Preliminary results indicate that these detectors outperform previously developed *ad hoc* detectors, thereby increasing the detection range and effectiveness of a proposed automated manatee avoidance system that is envisioned to reduce watercraft-manatee collisions.

5aUW8. Array gain of vector sensors in ocean noise. Andrew J. Poulsen and Arthur B. Baggeroer (MIT, 77 Massachusetts Ave., Cambridge, MA 02139, poulsen@mit.edu)

By measuring acoustic particle motion, vector sensors provide important capabilities to ocean sensing systems, such as the elimination of left/right ambiguity for towed arrays, the ability to "undersample" an acoustic wave without spatial aliasing, and improved detection performance and array gain in ocean noise fields. Several different ocean noise models exist, including isotropic noise, directional noise (both point and spatially spread directional noise sources), and realistic surface noise in a stratified ocean environment [W. A. Kuperman and F. Ingenito, J. Acoust. Soc. Am. **67**, 1988–1996 (1980)]. Theoretical expressions are derived for array data covariance matrices under different noise models, then used with optimal MVDR beamforming weights to analyze array gain. In order to better understand the advantages of vector sensor arrays, this paper also includes a performance comparison between vector and hydrophone arrays. Of further interest is array sensitivity and performance in the presence of system imperfections and mismatch between the assumed and actual array con-

figurations, or in the snapshot deficient regime. All arrays are susceptible to perturbations in array element locations, however vector sensor arrays are also sensitive to changes in sensor orientation. [Work supported by the PLUSNet program of the Office of Naval Research.]

10:15

5aUW9. Advantages of using diffracting cylinders and spheres in acoustic arrays. Philip W. Gillett, Marty Johnson, and James Carneal (Vib. and Acoust. Labs., Virginia Tech., 136 Durham Hall, Blacksburg, VA 24061, pgillett@vt.edu)

Acoustic arrays are designed using numerous geometries for different purposes. Line arrays provide for high beamformer gains in perpendicular directions, while circular and spherical arrays sacrifice those gains for the ability to discriminate between different sound source directions. Localization performance is limited to a specific frequency band governed by the geometry and spacing of the acoustic array elements. A different approach to acoustic array design is given involving the use of a rigid body as a mounting surface for the array elements. By placing elements directly onto a rigid object, the overall array performance is enhanced at lower frequencies, effectively producing a larger aperture than for a free field array of identical geometry. Evenly spacing elements around a cylinder allows nearly equal performance in all azimuth directions, enabling accurate bearing resolution on the plane of the array. Furthermore, placing elements around a rigid sphere produces an array that can provide both azimuth and elevation angles for incident sounds. To take advantage of potential performance increases, an inverse method approach is introduced that involves characterizing arrays in an anechoic environment and finding an inverse relationship to calculate bearing information from measured sound.

10:30

5aUW10. Localization and tracking of acoustic stimuli using millimetric artificial lateral line flow sensing arrays. Saunvit Pandya, Yingchen Yang, Chang Liu, and Douglas L. Jones (Univ. of Illinois at Urbana-Champaign, Urbana, IL)

We report on the successful development and implementation of acoustic dipole localization and tracking algorithms for millimeter-scale flow sensors arranged in an artificial lateral line configuration. Our previous work includes the development of an artificial lateral line system a novel, bioinspired flow sensing array. The artificial lateral line aims to mimic the lateral line sensory organ of many species of fish, an organ that allows the fish to perform near-field hydrodynamic imaging to navigate, capture prey, or align to a flow stream. Many of the stimuli that these fish sense can be modeled as near-field dipoles with widely understood acoustic signatures. In our previous work, we successfully developed an artificial lateral line using micro-flowsensors for applications such as dipole localization, wake tracking, and flow imaging. However, while the applications and the platform were both novel, potential uses were limited due to the size of the microsensors (and correspondingly the sensor arrays) as well as the limited distribution of the technology. In this paper, we show for the first time that a commercially available array can also be successfully used as an artificial lateral line for observing the above-mentioned phenomena, thereby enabling a new biomimetic millimeter-scale flow-sensing platform.

10:45

5aUW11. Active acoustic time reversal for underwater acoustic barriers. Sérgio M. Jesus (Inst. for Systems and Robotics, Univ. of Algarve, PT-8005-139 Faro, Portugal)

This work addresses the possibility of using successive transmissions of time delayed channel probe pulses between two closely spaced acoustic sensor arrays for forming an acoustic barrier for target detection in shallow water. One array is a transmit-receive array (TRA) while the other is a receive only vertical line array (VLA). The two arrays are connected via cable or wireless. Time reversed replicas of the acoustic channel response to the probe signals are retransmitted into the ocean propagation plane to form focus peaks at each VLA element. It is shown both theoretically and with simulated data that an optimum disturbance detector can be built from the data received at the VLA. This detector becomes sub-optimal due to usual time reversal drawbacks such as ocean non-stationarity and spatial sampling limitations. Real data tests are foreseen to take place during summer 2007 to answer questions such as allowable ranges and frequencies of operation [Work supported by FCT (Portugal) under project POCI/MAR/59008/2004].